30789

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* COLIN PARK

TITLE OF THESIS/*TITRE DE LA THÈSE* AN INVESTIGATION OF AN ASPECT OF REASONING

UNIVERSITY/*UNIVERSITÉ* U. of ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* PH.D.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1976.

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* CHARLES C. ANDERSON

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* 21st Oct. 1976 SIGNED/*SIGNÉ* C. Park

PERMANENT ADDRESS/*RÉSIDENCE FIXE* Faculty of Education
University of Otago,
DUNEDIN, New Zealand.

NL-91 (3-74)

INFORMATION TO USERS

THIS DISSERTATION HAS BEEN
MICROFILMED EXACTLY AS RECEIVED

This copy was produced from a microfiche copy of the original document. The quality of the copy is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Canadian Theses Division
Cataloguing Branch
National Library of Canada
Ottawa, Canada K1A ON4

AVIS AUX USAGERS

LA THESE A ETE MICROFILMEE
TELLE QUE NOUS L'AVONS RECUE

Cette copie a été faite à partir d'une microfiche du document original. La qualité de la copie dépend grandement de la qualité de la thèse soumise pour le microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

NOTA BENE: La qualité d'impression de certaines pages peut laisser à désirer. Microfilmee telle que nous l'avons reçue.

Division des thèses canadiennes
Direction du catalogage
Bibliothèque nationale du Canada
Ottawa, Canada K1A ON4

UNIVERSITY OF ALBERTA

AN INVESTIGATION OF AN ASPECT OF REASONING.

by

© COLIN PARK

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1976

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "An Investigation of an Aspect of Reasoning" submitted by Colin Park in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Charles C. Anderson
.....

Supervisor

.....

.....

.....

.....

Date October 21st, 1976

This thesis is respectfully, and properly,

and very happily dedicated

to Professor Charles C. Anderson.

ABSTRACT

Two problem-solving tasks much discussed in the literature recently, the 'selection task' and the 'rule-finding task', both initiated by P.C. Wason (1960, 1968), have been replicated, and the empirical findings of earlier workers are largely confirmed. However, it is suggested that claims regarding their importance are exaggerated, and in particular that the relation with Popper's notion of "refutability" in science is at best tenuous. The problems of a "scientific psychology" are briefly discussed, illustrated by the present area of investigation. It is proposed that in the field of science refutation can be meaningful only within an established conceptual structure, and that such a structure is lacking in psychology. Reasons are put forward as to why such a structure has not yet been established.

ACKNOWLEDGEMENTS

My sincere thanks are due to my advisor, Dr. C.C. Anderson, who has been unfailingly warm in his encouragement, and patient with my procrastinations. His clear-sighted criticism of much work in psychology has been a model for me throughout the last two years, and has only left me with the regret that I did not get to know him much earlier.

I owe a particular debt of gratitude to Professor Steve Hunka, not only for his work on the present thesis, but especially for his kind support and concern during the last four years in which I have worked in the Division of Educational Research Services, of which he is Coordinator. Like all who come into contact with him, I have been impressed with his integrity in academic and social matters, and his empathy in human relationships.

My thanks go to all my Committee members also, and to Dr. Jack Goldberg, for their criticisms and comments. My thanks also to the Principal and Grade 11 and 12 students of J.H. Picard School of the Edmonton Catholic School Board, and very particularly to Marcel Lavallee, Vice-principal, who very kindly and efficiently arranged the testing schedules, and encouraged students to take part in the experiment. My thanks also go to Dr. Alan Pearson of the

Department of Educational Foundations of the University of Alberta, for giving me time from one of his classes for testing, and for his very positive attitude to me and to his students regarding my work. I also have to thank all my subjects for giving me their time, and for their patience with my curious tasks.

Finally, my thanks to my wife and children for their support during the last four years.

TABLE OF CONTENTS

Chapter		Page
I	Background.	1
III	Experimental Procedures.	45
III	Results.	63
IV	Discussion	86
	Bibliography.	125
	Appendix I Elements of the propositional calculus.	133
	Appendix II Protocols from the selection task.	137
	Appendix III Protocols from the rule-finding task.	153
	Appendix IV Reasoning tests.	159

LIST OF TABLES

Table	Description	Page
1	Selection task: number of selections in each category; P.D.A.D. Students.	73
2	Selection task: number of selections in each category; B.Ed. Students.	74
3	Selection task: number of selections in each category; High School Students.	75
4	Selection task: Number of selections in each category; Computer Personnel.	76
5	Number of cases at each level of insight; P.D.A.D. Students.	77
6	Number of cases at each level of insight; B.Ed. Students.	78
7	Number of cases at each level of insight; High School Students.	79

LIST OF TABLES (Continued)

Table	Description	Page
8	Number of cases at each level of insight; Computer Personnel.	80
9	Numbers gaining insight after different choices; P.D.A.D. Students.	81
10	Numbers gaining insight after different choices; B.Ed. Students.	81
11	Numbers gaining insight after different choices; High School Group.	82
12	Numbers gaining insight after different choices; Computer Personnel.	82
13	Cell means and analysis of variance table, High School 'travel' problem.	83
14	Comparison of insight of science and non-science groups; abstract problem.	84

LIST OF TABLES (Continued)

Table	Description	Page
15	Comparison of insight of science and non-science groups; concrete problem.	84
16	Comparison of insight of four experimental groups; abstract problem.	85
17	Comparison of insight of four experimental groups; concrete problem.	85

LIST OF FIGURES

Figure	Description	Page
1	The abstract form of the selection task.	6
2	Wason and Johnson-Laird theoretical model for the selection task: preliminary model.	26
3	Wason and Johnson-Laird theoretical model for the selection task: revised model.	27
4	Choices according to the Wason and Johnson-Laird preliminary model.	28
5	The truth table for p implies q.	134
6	The truth table for p implies q and q implies p.	135

CHAPTER I

Background

Introduction.

In this thesis are reported two experiments, each involving essentially one aspect of reasoning. In the first, subjects are asked to select those cases of four given cases which would prove a given rule true or false. (This task will be referred to below as the "selection problem"). Many subjects fail this task, most in a fashion which can be included in one or other of a small number of error-types. In the second task, subjects are given an example of something which 'conforms to a rule the experimenter has in mind'. The subject is asked to suggest further examples, and is immediately told each time whether his example conforms to the rule or not. His task is to identify the rule. (This task will be referred to below as the "rule-finding problem"). Here, though few subjects fail completely, many show surprising inefficiencies in their approach to the problem. Both tasks will of course be described in detail below.

These experiments are not original, but are basically careful replications of earlier studies. These were initiated by P.C. Wason, the rule-finding task in 1960 (Wason, 1960), and the selection-problem in 1968 (Wason, 1968), and have been subsequently studied mainly by Wason and his coworkers and students, though more recently by a few apparently independent researchers. (The word "apparently" is

inserted here because it is known that at least one, Paolo Legrenzi, who published a paper based on his own work in the area in Italy, had earlier spent some time with Wason in London).

There exists, of course, a connection between the two tasks, and indeed it is the nature of this connection which provides the source of interest and motivation for the present work. This link is the failure to test for falsity, or to try to disprove a rule or hypothesis. In the selection problem, one possible choice which is critical, which could not be omitted as a test on any reasonable reading of the rule, and which could falsify the rule, is frequently omitted by mature, educated, adult subjects. In the rule-forming task, similar subjects often form a hypothesis, then "test" it by offering several confirming exemplars, instead of counter-examples. This failure to refute is said by earlier authors (see below) to be critically important on two counts: firstly, Popper has repeatedly emphasized the importance of refutability in science (eg. Popper, 1972), and its basic relevance to any objective "knowing" can scarcely

be reasonably questioned[1]; secondly, in Piaget's formulation of the stage of "formal operations" (which, by definition, is concerned with operations upon hypotheses), it is explicitly stated - and forms one of the elements of Piaget's basic and elegant INRC group - that, given a hypothesis of the form "if p then q", a person operating at this level will test the hypothesis specifically by looking for counter-examples, "x and not-y", (where x is an exemplar of p, and y is an exemplar of q). Since the point is crucial to any theory of a 'formal operational stage', it is as well to quote Piaget verbatim:

> In the special case of p implies q, p and not-q, q implies p, and not-p and not-p and q, the observer will note, for example, that faced with a complex causal situation the subject will ask himself two kinds of questions: (a) whether fact x implies fact y (which he himself will often express by two propositions which we shall call p and q and which he will link with the words "if

---

[1] Lakatos has pointed out that, in principle, one can never refute a hypothesis completely, any more than one can confirm it completely (eg., Lakatos & Musgrave, 1975, specifically p.184: "According to the Duhem-Quine thesis, any theory ... can be permanently saved from 'refutation' by some suitable adjustment in the background knowledge in which it is embedded"). However, it seems to me reasonable that in any one context, only a small part of a particular theory or hypothesis would normally be under question, and what is more, the assumptions on which the hypothesis is critically based can be generally agreed without much ado. Under these circumstances, refutability is a reasonable requirement for acceptance of any hypothesis. It could, of course, be claimed that I am invoking both Popper and Lakatos in an extremely naive fashion, and indeed I am. However, the situations dealt with in the present work are so straight-forward that it would be totally unreasonable to question the falsification requirement. Moreover, the application of the falsifiability requirement in Psychology could save the research community much time and effort - see Park(1972) and Platt(1964). See, however, the discussion in the last chapter below.

> (p) then (q)"). To verify it, he will look in this case to see whether or not there is a counter-example x and non-y, therefore p and not-q. (b) He will also ask whether it is really x which implies y ....(in Beth & Piaget, 1966, p. 181)

The status and authority of Piaget in modern psychology is unquestionable. The influence of his 'stage' theory of cognitive development likewise has dominated other theories to the point of eclipsing them - yet what happens to the theory if a 'stage' may vanish under the threat of unfamiliar material?

The reader who has not met the particular problems involved in the present work may feel reluctance in accepting the difficulty subjects have, and may suspect a 'catch'. This suspicion in fact provided the second reason for embarking on this study: Could such transparently simple problems present such difficulties to the cream of a population? - in the case of Wason and co-workers, to University students. In particular, could the experimenters have unwittingly but repeatedly made things in some way obscure or difficult for their subjects? (The possibility of continuity in this respect between experimenters explains the reference above to "apparently independent researchers").

These findings, in which intelligent adults fail repeatedly to solve a simple problem, and have great difficulty in attaining rational behaviour, even under guidance, clearly have implications for any educational activity above the level of inducing rote-learning. If intelligent and educated adults cannot reason logically with simple abstract

material, how can we expect immature children, groping in a mass of unfamiliar and uncoordinated new material to do any better? These deceptively-simple findings are therefore examined in some detail in this thesis.

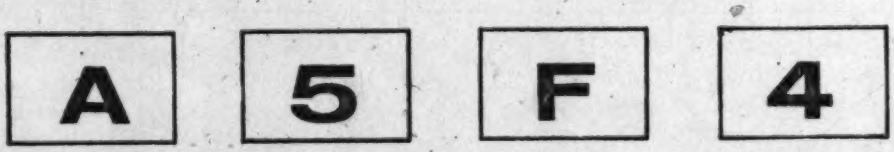
There are two further points to be made before proceeding with a review of previous work in the area. The first is that very clearly only a narrow aspect of reasoning is being considered here. It might reasonably be asked, How does this area of investigation tie in with the wealth of other data accrued over the last fifteen years or so in areas variously labelled as 'problem solving', 'information processing', 'concept formation', 'rule formation', and so on? The central point here is that <u>each</u> of these areas turns out to be rather narrowly defined in terms of its content, which fact is <u>not</u> obviously indicated by the titles attached to them. What communality is there, then, between the areas? Do we have <u>a</u> Science of Psychology, or several sciences, or any at all? These questions may well be closely tied to the astonishing variability and adaptability of human beings (with limitations within any one area), which perhaps <u>force</u> on experimenters these barely-connected areas in which to work, and may account at a very fundamental level for basic difficulties in establishing a scientific psychology. Possibilities arising from such thoughts will be discussed in the last chapter, and there will be no review of work in such 'wider' aspects of reasoning in the present chapter.

The second point concerns the comprehensibility of the material below to readers unfamiliar with experiments in this area. One form of each of the "selection problem" and the "rule-finding problem" will be described here so that it will be simpler to follow the description of and comments on previous work. For completeness, these descriptions will be repeated in the next chapter.

Brief description of the tasks.

(1) The Selection Problem.

In one form of this task, four cards are presented, each of which has some characteristic showing (eg. a blue triangle, or a vowel, or a border of a figure) and some characteristic hidden (on the reverse side, or by an opaque mask). The subject is given a rule, such as, "Every card with a vowel on one side has an even number on the other side".



The abstract form of the selection task

Figure 1

The subject is asked to select those cards he would like to look at to be sure that the rule is true or false for all four cases. The correct choice for the case shown in Figure 1 would be the 'A' and the '5'.

(2) The rule-finding problem.

Two forms of the problem have been used, one including sets of three numbers, the other involving animals. In the former case, the subject is told "I am going to give you an example of something conforming to a rule I have in mind. I would like you to give me examples, one at a time, and I will tell you whether each one conforms to my rule, or does not conform to my rule. The first example is '2,4,6' ". The rule is, in fact, any set of three numbers in ascending order, (which, of course, does not exclude odd numbers, negative numbers, zero, rational, irrational, or transcendental numbers, nor the possibility of unequal increments).

Previous work on the selection task.

In his 1968 paper, "Reasoning About a Rule", Wason reports that earlier work at the Psycholinguistics Research Unit of London University had shown that in a task involving propositions expressed in everyday terms, affirmation of the consequent occurs significantly more often than denial of the antecedent (Wason, 1968(b), p. 273). This was also the case with abstract material, but to a greater degree: Labelling the antecedent and consequent p and q respectively, (eg. if there is a vowel on one side of a card (p) then there is an even number on the other side(q)), he reports:

> Nearly all subjects select p, from 60 to 75% select q, only a minority select not-q, and hardly any select not-p. Thus two errors are committed: the consequent is fallaciously affirmed, and the contrapositive is witheld[2] (Wason, 1968, pp. 273-274).

On the grounds that the semantic concept of falsity and the syntactic concept of negation "both cause difficulty when sentences have to be evaluated or constructed", Wason therefore attempted 'therapeutic' treatments, to "correct the bias towards truth or correspondence" (ibid, p. 274). In the first experiment, subjects performed the selection task (with four cards, using letters and numbers, and the rule "If there is a 'D' on one side of any card then there is a '3' on its other side"), and were then asked, for each card in turn, what values on the backs of the cards would make the rule false. As Wason puts it, they were invited

---

[2] See Appendix I for an outline of logical terms.

"to project falsity". They were then invited to revise their first decision. A control group was simply asked to think again about their decision "because people often do this task too quickly, and get it wrong". (Wason comments in passing that "two subjects ... seemed unable to comply with the instructions and were hence rejected". This point will be discussed later). The efficacy of Wason's therapy was not great: the frequency of selecting not-q increased from 5 to 8 (out of 18) in the experimental group, from 2 to 3 (out of 16) in the control group (after 'reconsidering'). More pertinently, of those experimental subjects not selecting not-q initially (13) 8 'projected' p on to not-q, but only 3 of these 8 included not-q in their revised selection. "Thus", says Wason, "the therapy of falsifying values cannot always be induced, and even when it is induced it is by no means effective" (ibid, p. 275).

A further point which will again be discussed below, is a somewhat curious comment made by Wason regarding the responses of some subjects when asked to pick out only those values which 'could break the rule', ie. falsify the conditional sentence. Four subjects selected only values of not-p and not-q, and refused to turn them over, because they claimed this was useless. He quotes, "It doesn't make any difference - the two I have chosen do break the rule", and "There is no rule regarding that card (not-p)". Wason remarks,

> Thus, in a small minority of subjects, the concept of something following a rule appears to be inade-

> quately conceived, for to know what could follow a rule is to know what could break that rule (Wason, 1968, p. 277).

In the second 'therapeutic' attempt, four "program cards" were prepared with the following stimuli on either side: (a) square, yellow scribble; (b) square, red scribble; (c) rectangle, red scribble; (d) hexagon, brown scribble. The experimental group was then presented with the sentence, "If there is a square on one side of the card, then there is a red scribble on the other side", and were asked to pick out from the four program cards "the one card which makes the rule false", and then "any which make the rule true". (That is, they could handle and examine the cards). It was explained to them that "their decision meant that the converse of the sentence could not be assumed - that the rule held 'only one way'" (ibid, p. 278).

Then both experimental and control groups were presented with all eight combinations of p and not-p, and q and not-q, and were asked to select those cards which would show "if you knew what was on the other side", that the rule was true or false. They were then asked to turn over the cards they had indicated, and to tell the experimenter whether each proved the rule true or false. Finally, subjects were asked to "project falsity" on to the cards they had not selected: starting with the not-q cards, they were asked, "Could anything on the back of this make the rule false?"

The "programming procedure" produced little improvement. Though all subjects in the experimental group picked

p and not-q as the only falsifying case, and p and q as the only verifying case during the programming, the responses of the two groups were almost identical in the selection task; in particular, one person out of thirteen in each group chose the case of p and not-q.

In his discussion of this paper, Wason is therefore led to question whether the stage of formal operations is completely achieved at adolescence "even among intelligent individuals". He further asks whether it is the verbal trappings of the problem which create the difficulty, or is the difficulty inherent in the 'formal structure' of the rules? And if the words in which the rules are expressed are responsible for the difficulties, "what words illuminate the structure?" (ibid, p. 281). As I hope to make clear in the sequel, I believe the latter formulation is a red herring - I do not believe that "the rules" and "the words in which they are expressed" are psychologically separable.

Subsequent investigations by the Wason and Johnson-Laird group.

Having raised the question of whether the wording of the problem might affect subjects' performance, the effect of presenting the disjunctive form of the rule was next tested (Wason & Johnson-Laird, 1969). The authors point out that the truth table for disjunction, that is, "either there is not-p or there is q, or both" (not-p or q) and the truth table for implication (p implies q) are identical. However,

the work described above (Wason, 1968) and an unpublished work by Johnson-Laird and Tagart show that "if p then q" is usually understood to be true only when both p and q are true. (This was in fact confirmed during the present investigation).

A modification was made in the 1969 paper, in that subjects were not asked to pick out all the cases to prove the rule true or false, but were told in one alternative that the rule was true, in another that it was false, and were told to choose values to prove this. (The correct response is of course the same in both cases). Two forms of the sentences were used, corresponding to the disjunctions (not-p or q) and (p or q). In the first case, the correct response is to select not-p and not-q, and in the second case, to select p and not-q. Two such forms were used because of the known difficulty in evaluating negative sentences: it was hoped that the (p or q) form would "act as a baseline against which the difficulty of (not-p or q) could be measured" (Wason & Johnson-Laird, 1969, p. 15)[3]. Four different stimulus-sentences were used with each of these con-

---

[3] It is, however, difficult to see exactly to what use the authors could put such a comparison. The aim is to compare the difficulty of (not-p or q) found in this work with the equivalent implication form (if p then q), reported in Wason, 1968. Presumably, what is intended here is that any extra difficulty which might be encountered could be blamed on the negative form of the proposition ("either there isn't a p or there is a q"), and therefore the non-negated form must be brought in for comparison. But this (p or q) is a logically and linguistically <u>different</u> proposition (in fact equivalent to the implication "if not-p then q"), and so cannot be directly compared in difficulty with "if p then q".

ditions - eight sentences in all. For each of these, eight stimulus-cards were constructed, consisting of exemplars of p or not-p on one side, and q or not-q on the other side.

Each subject was presented, one at a time, with four different sentences, each exemplifying one of the four different conditions, ((p or q)T; (p or q)F; (not-p or q)T; and (not-p or q)F), where the T or F after the conditions indicates of course that the subject was told that the rule was true or false, respectively, for that case.

For each condition, subjects were given scores of 0 to 4, according to the number of correct cards he had selected. Mean scores were:

(p or q)T = 2.9; (p or q)F = 3.3;

(not-p or q)T = 2.6; and (not-p or q)F = 2.7.

The prediction that the negative conditions would be more difficult than the affirmative ones was confirmed, but there was no significant difference statistically between the true and false conditions. For both true and false cases, correct selections were made more frequently than incorrect ones - 36 out of 48 for the case (p or q) and 25 out of 48 for the case (not-p or q), but only four subjects out of twenty-four were correct for all four conditions.

Though the task in the form "either not-p or q" thus appears to be easier than that outlined above for "if p then q", (summarizing the Wason, 1968 paper), once again, more interesting aspects arise from the qualitative observations. Thus, when the rule was known to be true, four subjects

"matched" on both 'true' conditions (i.e. chose those cards whose exposed sides were mentioned in the rule - which is of course an incorrect strategy), yet three of them made correct selections under the 'false' condition. When the rule was true, reasons given for rejecting the remaining cards were:

"You are trying to prove the rule is true, and these might disprove it" [sic!].

"These others are not what is stated in the rule, and the rule is true".

"I chose the true ones. No matter what is on the other side, these would proved [sic] the rule" (Wason and Johnson-Laird, 1969, p. 19).

However, when the rule is false, these same subjects made the correct selections for the wrong reasons:

"If I am trying to prove the rule false, then it is obvious you wouldn't start with the ones which prove it correct".

"Those others are what is stated in the rule and the rule is supposed to be false" (ibid).

However, beyond these particular cases, 16 of the 24 subjects spontaneously doubted that some of their selections were correct. As Wason and Johnson-Laird put it,

> "With a conditional, the individual is likely to be confident but wrong; with a disjunction, he is likely to be unconfident but right. The meaning of the conditional gives no hint of the negation or falsity which underlies its logic. The disjunctive element makes this element explicit, but this seems to weaken the grounds upon which any inference can be made"(ibid, p. 20).

Wason attempted to remove the "fixation" that the case of not-q is irrelevant in a paper provocatively entitled "Regression in Reasoning" (Wason, 1969a). The rule the subjects were to test was the sentence :

"Every card which has a red triangle on one side
has a blue circle on the other side.",

and the visual stimuli were cards with a red triangle (blue circle); a red circle (red triangle); a blue triangle (red circle); and a blue circle (red triangle) - where the symbol named in parentheses was the hidden one of each pair. Subjects were told that each card contained a circle on one side and a triangle on the other, and were invited to examine both sides of each of a set of eight training cards, four with circles facing up, four with triangles up. The cards were then replaced with the four test cards listed above, the rule was stated, and subjects were asked to tell the experimenter "which of the cards you need to turn over in order to find out whether the sentence in front of you is true or false". If the subject said only that one showing the red triangle, the experimenter asked 'in a casual voice', "Only the red triangle?" The subject was then asked what there could be on the reverse of the red triangle card. If he said "a blue circle", he was reminded that the rule <u>could</u> be true or false. If the subject did not arrive at the conclusion that the p and not-q condition would prove the rule false, he was told that it would. He was then asked, "By the way, what was your choice of cards to turn

over in order to find out whether the rule in front of you is true or false?" In this way, the subject had the chance of adding the case not-q(p). This sequence Wason called the 'weak hypothetical contradiction test'.

In the 'strong hypothetical contradiction test', the experimenter pointed to the not-q case, and repeated the above procedure. If the subject failed to say that the not-q(p) contingency would make the sentence false, he was prompted.

In the 'concrete contradiction test', the subject turned over those cards he had selected as relevant, and was asked if the sentence was true or false. If he had not chosen not-q, the experimenter turned it over - (this was in fact the only refuting case) - and the question was repeated.

With a separate group of subjects (each had sixteen members), the above experiment was carried out, but with training cards which had the contrast red or blue on the two sides, rather than triangle or circle, so that the subjects knew that the red circle (the previous falsifying case) could not have a red triangle on the other side, and the falsifying card in this case was thus the blue triangle. Wason predicted:

> ...When the colour is unpredictable by the subject, [the first situation above] performance would be better than when the shape is unpredictable. The stimulus which is 'not a blue circle', and hence would falsify the test sentence if it were on the same card as a red triangle, seems intuitively to be more appropriately satisfied by a red circle than a blue triangle. A red circle

> lacks only an attribute (the colour) of the blue circle, but a blue triangle is a totally different entity (Wason, 1969, p. 472).

The present author is unable to follow this 'intuitive reasoning', but in any event, there was no difference between the performances of the two groups, so this section of the experiment has not been treated in any detail here.

Confronting the subjects with this sequence of contradictions did in fact result in a monotonic, but irregular increase in the number of subjects selecting the critical not-q case. However, it was not until the concrete test that a significant proportion began to realize that only p and not-q need be turned over: after the 'strong hypothetical contradiction' stage, only five had decided on p and not-q. After the 'concrete contradiction', 10 now selected p and not-q, and the number selecting only p and q had dropped from 10 at the previous stage to zero. However, the number now selecting p, q, and not-q increased from 10 (at the 'strong hypothetical test' stage) to 17 at the 'concrete' stage. Subjects certainly showed reluctance to reach Wason's Received Truth of p and not-q. As Wason says, "These modifications were designed to make the problem easier. They did not succeed in doing so" (ibid, p. 476).

Wason points out that his method could be criticized on the grounds that (i) The hypothetical contradictions were asked about contingencies other than the falsifying ones, p(not-q) and not-q(p), and could thus have been inefficient; and (ii) that these tests "might have occurred too far in

time from the subject's previous choice of cards".

To counter these arguments, a further small study was carried out in which nine subjects were asked only about the consequences of the two falsifying contingencies, and were immediately asked to recall the task; they were then asked to recall their previous choice, and whether they were satisfied with their choice.

Wason reports that these subjects did worse than those in the main experiment, refuting the criticisms with reasonable certainty. However - and this is the point of mentioning the subsidiary experiment here - Wason offers a possible 'explanation', showing why this treatment might have caused a worse performance, because

> subjects may have been defending their previous decisions unconsciously against contradictory information. The attempt to strengthen the contradictions may have increased the momentum of the original intentions so that they remained immune from critical information self-generated by the subjects. The proximity of this critical information to the subject's decision may, paradoxically, have insulated the decisions against interference (ibid, pp 477-478).

In other words, Wason provides a possible mechanism to explain a certain phenomenon, and an experiment to test his explanation. The experiment turns out to have results opposite to those expected, and the strengthening of 'unconscious defenses' is invoked to explain this. But presumably, if the experiment had had the expected results, the belief that it was the true explanation would have been strengthened. Its refutability has been shielded by the introduction of other factors, untested, and possibly untes-

table. This kind of sloppy thinking regarding 'explanations' is common in psychological papers[4],[5] but is particularly regrettable in a paper devoted to logical thinking and the refutation of hypotheses[6]. To quote Wason, from this same paper:

> Piaget's view that the adolescent will seek to verify a 'complex causal situation' by looking for the counter-example, (p and not-q) in order to verify that p implies q (Beth and Piaget, 1966) is not corroborated by the data (ibid, p. 478).

To explain the failure in the experimental task of "highly intelligent subjects" Wason proposes that either Piaget's theory requires modification, or that "there is something about the task which predisposes some individuals to regress temporarily to earlier modes of cognitive functioning" (ibid, p. 478). But how are we to account for Wason's failure to 'look for the counter-example'?

Nevertheless, one clearly cannot disagree with the conclusion that the subjects, who should all be completely

---

[4] See Park, 1972, for other examples.

[5] However, matters are rarely clear-cut. Some weeks after writing the above footnote, I came across the following comment in a 'Reply' to Popper, from a former student and colleague of his:
"Although one gathers the impression that Popper takes a verificationalist, rather than falsificationalist, approach to the examination of his own intellectual development..." (Bartley, 1968, p. 116.) There are indeed many philosophers who think that Popper's theories are quite empty - see, for example, Schilpp(1974), throughout.

[6] It is not suggested, of course, that the experimenter should be determined to refute something every time he opens his laboratory door. As Medawar has said, "A realistic methodology must be one that allows for repair as readily as for refutation" (Medawar, 1969, p. 41). However, Wason is here explaining away his own explanation of a particular observation.

within the stage of formal operations, according to the Piagetian theory, failed to think in hypothetical terms (that is, to consider what might be on the other side of the stimulus cards), and to look for a counter-example. The wording of the rule in the paper considered here, and the use of only binary stimuli (red/blue, triangle/circle) did not apparently ease the problem for the subjects. The presentations of contradictions did indeed - finally - cause most subjects to admit the relevance of the card showing the not-q case, but seemed to cause more subjects (than in Wason's 1968 paper, outlined above) to select the q case as well. If one is thinking in Piagetian terms, the subjects do appear to be regressing.

Following this convincing demonstration of failure to reason adequately about conditional sentences in an abstract task, a report was published of an attempt to investigate

> (i) Whether insight would be more readily gained if the task were to be both concrete and simplified, and (ii) whether such insight is an all-or-none matter, or whether it would vary over a series of trials as a function of the cognitive load imposed (Johnson-Laird & Wason, 1970).

There seems to be a slight gap here, as far as the second investigation is concerned. Just what form of cognitive load is being considered, in relation to the earlier experiments? After all, we all know that we do not perform as well under greater 'cognitive load' (that is, presumably, with more difficult tasks) as we do with easier tasks. Wason had, furthermore, already produced evidence that 'insight', in the context of the selection tasks, is not an

"all-or-none matter". It seems likely, then, that the reason quoted above for carrying out the second experiment is actually an after-the-fact rationalization, and that the initial motivation was much less clear-cut[7]

Neither of the two investigations produced results in any way surprising when viewed in the light of the earlier work, and they will be desribed as briefly as possible here. The first experiment was designed "primarily to examine the grasp of implication in a simplified, concrete task". In addition, the two syntactic forms, "if they are p's, then they are q" (conditional) and "all p's are q" (quantified) were compared, and the semantic effect was tested of having subjects prove the rule true, or prove it false[8]. (As reported above, the latter factor was found to have no effect when implication was expressed as a negated disjunction, "either not-p or q"). One of four groups was assigned to

---

[7] This is not a condemnation of the experiment itself. Empirical science may necessarily be a fuzzy, stumbling sort of process, gaining its apparent clarity and sense of direction only on the published page. See P.B. Medawar, Is the Scientific Paper a Fraud? (1963). The problem - as here - is that the development of an idea may be lost to the reader. This is well-illustrated in the educational sphere in the 'purest' of the sciences, physics and chemistry, when high school students want to know why someone was studying a particular problem at all. What led Boyle and Charles, for example, to "mess about with gases" (as one student put it)? Or Rutherford to bombard nitrogen with the emanations of radium? The answers are not easy to find in published experimental reports, and indeed, Medawar claims that we should rarely expect to find them: "It is no use looking to scientific 'papers', for they not merely conceal but actively misrepresent the reasoning that goes into the work they describe" (Medawar, 1967, p. 151).

[8] Johnson-Laird and Wason used "p" and "q" here to refer to particular exemplars, not to propositions. Hence, they are not underlined.

each of the conditions formed by combining these two pairs, (seven subjects per group). The test sentences used were,

All the triangles are black.

and

If they are triangles, then they are black.

The sentences referred to the contents of two boxes, one labelled "white" and one labelled "black". Subjects were assured that the contents were indeed white and black respectively, and that only circular and triangular shapes were contained in the boxes. They were to ask for shapes one at a time from either box, and announce when they were sure the rule was definitely true or false. Each box contained fifteen shapes, and of course there was need to ask for white objects only, and to select them all. For subjects proving the rule false, the white box contained fourteen circles and one triangle. In order to give subjects the chance of demonstrating spontaneous insight, if they requested black shapes, they were given only triangles on the first five trials, and a circle on the sixth. In the two groups proving the rule false, all fourteen white circles were presented before the white triangle, so that insight was revealed by persistence in requesting only white stimuli.

Essentially, the task was found to be easy. No subject asked for all the black shapes, and all of them requested all the white ones, so that all were assumed to have gained insight at some point. The total number of black shapes requested was hence used as an indication of insight (or

rather, of lack of insight). There was no statistically significant difference between the 'quantified' and 'conditional' cases, but the 'prove false' condition produced significantly fewer errors than the 'prove true' condition (using a Kruskal Wallis analysis of variance, $P < 0.01$). This is to be expected in this simple case, since in a sense, "show the rule false" is directing subjects to the white cases. Indeed, Johnson-Laird and Wason report that only one subject requested six black stimuli, and none of the others requested more than two. (In fact, nine out of the fourteen in the 'prove false' condition requested no black shapes at all).

The "gradually increasing cognitive load" in the second experiment was derived from the use of six sentences involving the quantifiers "every", "some" and "no", used as follows (from Johnson-Laird and Wason, 1970, p. 54):

(i) Doubly-quantified rules:

(a) Every dot is connected to some dot or other.

(b) No dot is connected to every dot.

(ii) Conjunction or disjunction of doubly-quantified rules.

(c) There is a dot which is connected to a dot, but no dot is connected to every dot.

(d) There is a dot which is not connected to any dot or every dot is connected to every dot.

(iii) Triply-quantified rules.

(e) There is a dot connected to a dot to which no other

dot is connected.

(f) Every dot is connected to a dot to which another dot is connected.

It was assumed that these rules increased in complexity over the three levels. Hence, it was predicted that the simpler the rule, the greater the insight which would be gained into the structure of the task. Such insight would be manifested in the absence of positive instances chosen by the subjects and the presence of negative instances.

Each of 24 subjects was tested in all six rules in a counterbalanced order.

The task was this: Subjects had in front of them eleven cards, each with four dots in the form of a square. Straight lines connected none, some or all of the dots on each card, and no two diagrams were alike. The instructions from the experimenter were

> I want you to imagine that I have taken some of these diagrams and put them in an envelope, sealed it, and then written a description of all the diagrams it contains. Of course, I haven't put all the diagrams in the envelope, and the description might also apply to some of the diagrams left outside the envelope. Your task is to discover whether my description of the contents of the envelope is true or false. The way you will do this is by picking out, one at a time, those diagrams which you want information about. I will tell you whether each diagram you choose is inside or outside the envelope.

A diagram which is a positive instance of the rule may be inside or outside, and knowledge of its location is of no use to the subject. Hence, to check the truth or falsity of

the rule, he must ask for information on non-exemplars. As subjects selected diagrams, they were asked to state whether their choice was a positive or a negative instance. (In one of the subsequent analyses, Johnson-Laird and Wason counted as a 'correct' choice one in which the subject pointed to a positive case, but identified it as negative - in other words, he had intended to choose a negative exemplar).

Five of the 24 subjects showed no insight, choosing both positive and negative instances throughout, and four showed complete insight; data from all nine were discarded as "unilluminating"[9]. Data from the remaining 15 subjects essentially confirmed the predicted decrease in insight with complexity of rule, but the expected trend was by no means clear-cut. The chief cause of this lack of clarity was the difficulty encountered by subjects with the fourth rule, the disjunction, "There is a dot which is not connected to any dot, or every dot is connected to every dot". The difficulty was not caused by misidentification of instances of the rule; on the contrary, together with rule (a) it had the lowest rate of misidentification of all - 1.5% for both rules. According to the linguistic intuition of the present author, the rule (d) seems to be written in a particularly

---

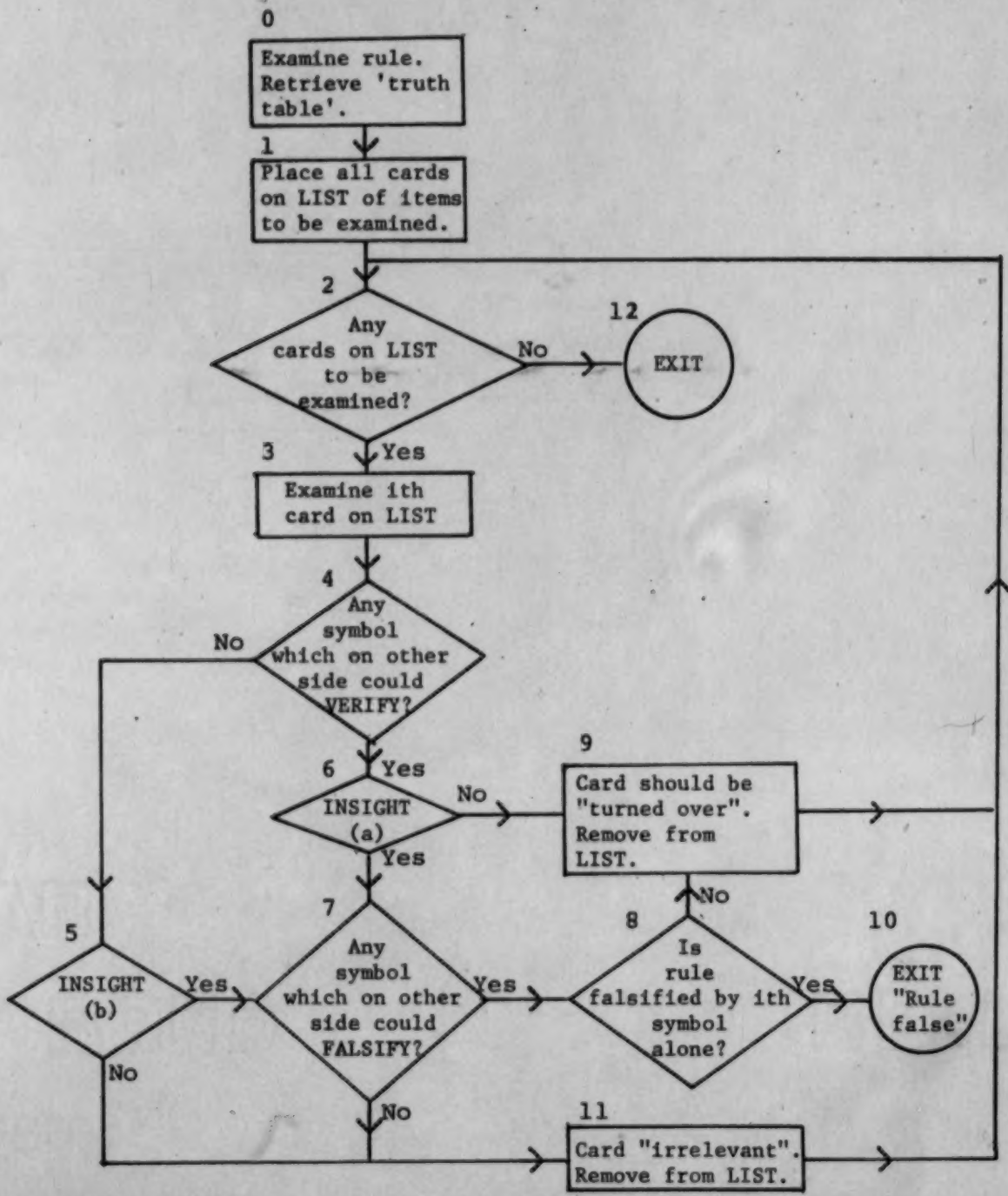
[9] It seems regrettable that the data from the five showing no insight were discarded. Did they perform relatively better with the 'easier' rules? Were they beginning to show a tendency to choose negative examples? (The authors report for the remaining 15 "a slight but insignificant tendency for insight to increase with practice" (ibid, p. 56). Did these subjects come from among those who received the more difficult rules first?

obtuse way. Though the remainder are hardly paradigms of good style, they seem to permit images to come to mind more easily than does rule (d). The difficulty of comprehension is further indicated by the mean inspection time during which each rule was examined before the first selection. These were, in seconds: (a) 29; (b) 40; (c) 37; (d) 55; (e) 41; (f) 53.

The results of all the above-mentioned experiments were subjected to a 'theoretical analysis' in Johnson-Laird and Wason (1970(b)). This 'analysis' consisted of the construction of two 'information-processing' models (flow-charts), a preliminary and a revised model. These flow-charts are reproduced on the following two pages.

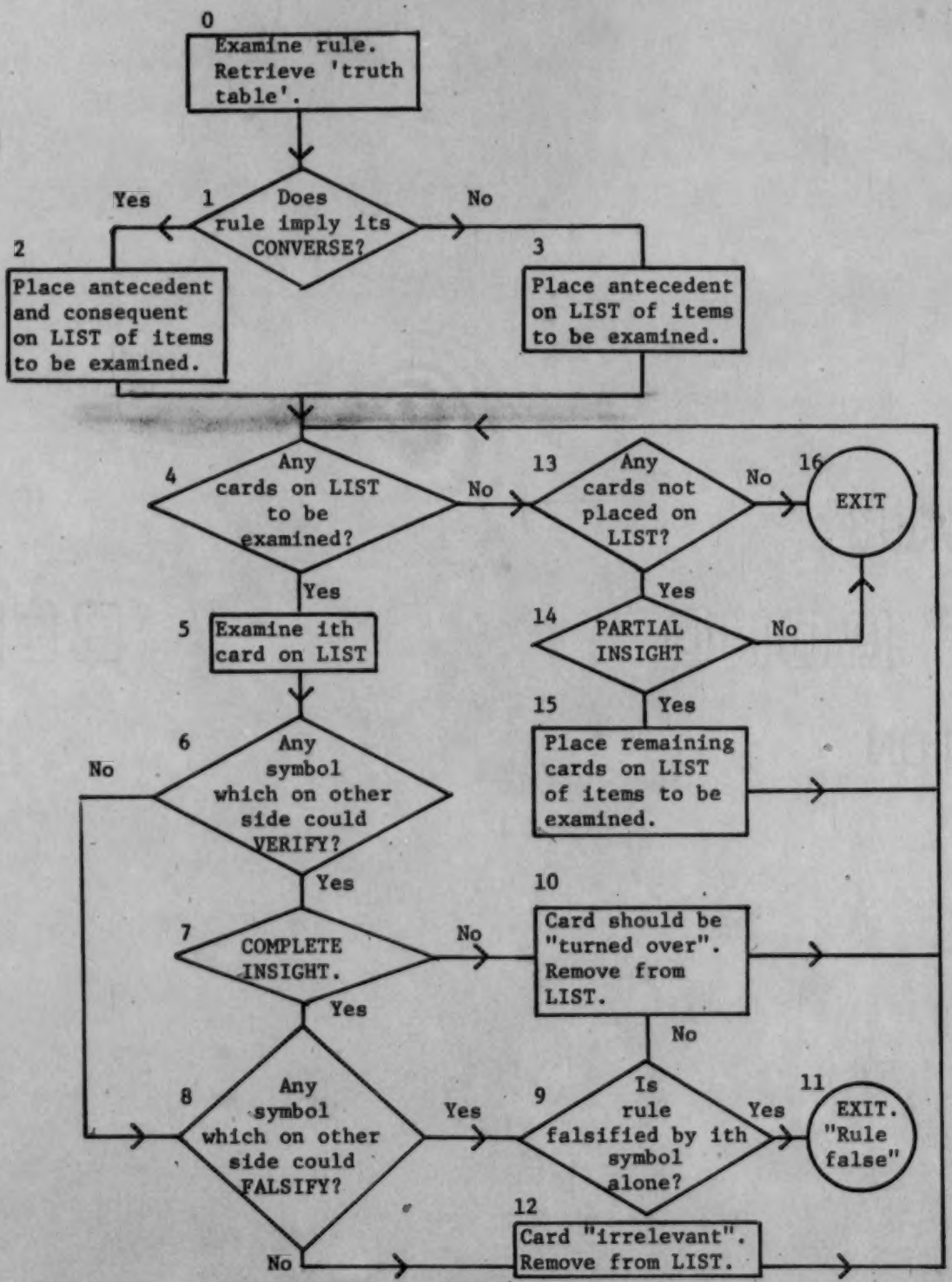
The first model seems to be rather obscure. Firstly, I take the notation in boxes 4 and 7 to mean "Could any symbol on the other side of this card VERIFY/FALSIFY the rule?", respectively.

Secondly, in their text, the authors do not in fact state explicitly what insights (a) and (b) are. A little spadework produces the following synthesis: Insight (a) demands that subjects ask the question, "Of those cards which could verify the rule, (i.e. p and q), could any symbol on the reverse <u>falsify</u> the rule?" If the subject does not have this insight, both the verifying cards are turned over. If he does, q is not considered further. Insight (b) demands that subjects ask the question, "Of those cards which could not verify the rule, (not-p and not-q), could any symbol on



Johnson-Laird and Wason (1970), "Preliminary model".

Figure 2.



Johnson-Laird and Wason (1970), "Revised model".

Figure 3.

the reverse falsify the rule?" If the subject does not have this insight, both cards are ignored as irrelevant. Thus, the presence or absence of each of the two insights give rise to the following selections of elements-to-be-examined:

		INSIGHT (b)	
		Present	Absent
INSIGHT (a)	Present	p, not-q	p
	Absent	p, q, not-q	p, q

Choices according to the Wason and Johnson-Laird "Preliminary model"

Figure 4

This model was rejected for three reasons: firstly, it seemed unlikely that insight (a) would occur three times as often as insight (b) (as was indicated by the data from Wason's earlier experiments), since the two seem conceptually similar. Insight (a) involves the rejection of a value because it could not falsify; insight (b) involves the acceptance of a value because it could falsify. Secondly, the selection of only p (insight (a) without insight (b)) "may not indicate a deep insight, but merely signify that S construes the rule as asymmetric" (op cit, p. 140). Thirdly, in the "remedial" treatments, subjects often relinquished the choice of p only, in favour of p, q, and not-q, which,

as is indicated in the table above, involves the gain of insight (b), with the simultaneous loss of insight (a) - which seems unlikely. Moreover, Johnson-Laird and Wason report (ibid, p. 140), that very few subjects exhibit the transition from p and q to p alone, which would correspond to the gain of insight (a), from the state of no insight.

The 'Revised Model' seems intuitively more satisfying. Firstly, most subjects select either p only, or p and q as their initial selection - 42 and 59 respectively, out of 128 subjects in the Johnson-Laird and Wason paper - (1970b, p. 136). These subjects are focussing only on cards mentioned in the rule. Beyond this point, two levels of insight are assumed, but "complete insight" entails "partial insight". The latter consists in recognizing that cards which falsify should be selected. Complete insight consists in recognizing that <u>only</u> cards which could falsify should be selected. (Note that the model does <u>not</u> necessarily explain the immediate selection of p and not-q by a few subjects, since we surely cannot maintain in the total absence of supporting evidence that they must have gone through the stage of partial insight first).

Johnson-Laird and Wason suggest (ibid, p. 144) that it is possible to test this model by examining, for example, reaction times, or by using other logical connectives. In my opinion, this is unlikely, but will not be discussed further; in particular, this seems to me to be pre-eminently the kind of experiment in which reaction-times should <u>not</u> be

used.

In summarizing, the authors once again make the point that the Piagetian notion of the formal-operational stage seems to be challenged by their results. They further add a novel remark concerning the maintenance of <u>prejudice</u>:

> ...It is plausible to suppose that certain prejudices are maintained in the face of contrary evidence because the prejudiced individual lacks a type of insight which is analogous to the insight required to solve our problem. A person who believes, say, that all actors are effeminate, is unlikely to test his belief by scrutinizing actors, because there is no immediate way of identifying them. Nor will such a person consider people who are not effeminate as relevant to his belief. What is more likely is that he will note the occupation of any effeminate individual he may encounter. If such an individual turns out to be an actor, the belief is confirmed. If he turns out not to be an actor, the belief is obviously (and quite validly) unaffected. In this way the prejudice is proof against falsification (ibid, p. 146).

This is an interesting and entertaining application of the authors' ideas, but is almost certainly irrelevant, for surely a common characteristic of most prejudices is that the objects of the prejudice <u>are easily identifiable</u>, by skin colour, by dress, manner, accent, and so on, in contrast to the "actors" of Johnson-Laird and Wason's example. In fact one imagines that it might have been quite difficult for the authors to construct their example initially, for the first examples to come to mind would be expected to be the visible (as well as well-known) examples.

The model described above was tested in part in Goodwin and Wason (1972). In this experiment, three variants were introduced: firstly, stimuli were positioned on the two

halves of the same side of each card, one half being covered (avoiding the difficulty that subjects might understand "the other side of the card" to mean "the side facing down", and avoiding the use of "a complex set of masks"); secondly, subjects were asked to write out, the reasons for their choices; thirdly, subjects in an experimental group had a full set of cards available to them to examine while they were working on the experimental task. A control group did not have this aid.

It appeared that the difficulty of the task was not due to the difficulty of remembering and manipulating the abstract symbols, for the presence of the extra set of cards did not improve the performance of the experimental group over the control group. Thus, the authors once more attributed the difficulty to "the failure of the abstract material to induce the idea of a combinatorial analysis" (ibid, p. 206). They also stated that the reasons given by subjects for their choices corroborated the information-processing model above, in that three levels of insight were shown, and "there is a close relation between the degrees of insight indicated by the protocols and the corresponding selection of the cards (ibid, p. 212). However, the process of giving reasons did not apparently help in producing insight:

> Only five out of the 26 subjects attained the correct solution after trying to give reasons for their incorrect selections, and two of them had already attained partial insight. On the other hand, two subjects regressed, from partial insight, and complete insight respectively, to a

> total lack of insight, revealed by changing their selections to p and q.

Two points of interest arise from this paper. Firstly, six subjects out of thirty-two were classified as "other" in their selection of logical values. (The other tabled choices were: p and not-q; p, q, and not-q; p and q; and p alone). Five of the protocols of these six were not presented "because most of them are infected [sic!] with serious confusion and apparent misunderstanding of the problem" (ibid, p. 211). That is, almost 20% of the "intelligent subjects" failed to grasp the problem! Secondly, two of these six chose q only, which Goodwin and Wason attribute to an extended verification set:

> This rare selection evidently arises when the set for truth extends to a presupposition of the truth of the claim to be tested (p. 208).

However, it is surely crucial to point out that the verification set is working <u>in one direction only</u>: the subject assumes that a p will have a q with it - (in the one protocol given which involves the choice of q only, the subject <u>says</u> so) - but he does not then assume the q must have a p with it. This is a classic example of the irreversibility of which Wason makes so much, and which he relates to the Piagetian framework. (The subject has apparently 'projected' a q on to the p, then, assuming the rule true, decides not to reveal this case. But what has the subject projected on to q? For if it is a p, he need not test it, and if it is a not-p, he also need not test it, <u>as he has already shown</u> in his rejection of the not-p case).

Of the two subjects choosing q only, Goodwin and Wason say in their Discussion section, "...these two subjects appear to be more rational [than earlier subjects who had shown a verification set by choosing p and q], because, if they assume q is associated with p, there is no need to select p" (ibid, p. 211). This somewhat confusing argument, though possibly legitimate, is confounded by a misprint in the same section; the above-quoted sentence is preceded by "One novel feature of the present results is the selection of only not-q (two cases)" (ibid). This must surely refer to the two q-only cases just discussed, as is further indicated by the similarity of the language of the paragraph in question (in the Discussion section) and that under "Selection q" in the "Qualitative Results" section (pp 209-210). Substituting the negation of a predicate for the predicate would be a grievous error in any context!

In a paper published slightly earlier than that just discussed (which was introduced first because of its closer association with the information-processing model), Wason and Shapiro investigated the effect of having two groups of subjects 'project' what values on the reverse of each of 24 cards would make a given rule true or false (group 1), or say whether the value in fact on the reverse of each card made the rule true or false (group 2) (Wason and Shapiro, 1971). They were told, however, that it would be in order to say that no value on the other side would make the rule true or false. Both groups were corrected when they made

errors. The authors hypothesized that since "the construction method clearly involves an imaginative act, and hence a greater degree of involvement, ... it was predicted that it would be associated with superior performance in the subsequent selection task" (ibid, p. 64).

The selection task was in fact presented to all subjects in two forms. One used letter and number stimuli identical to those in the training tasks, the other, the "transfer" task, used different letters and numbers, along with a different rule. The first rule was "Every card which has a vowel on one side has an even number on the other side"; the second, "Every card which has a D on one side has a 3 on the other side".

The experiment showed "a trend in favour of the construction group", but the difference was not statistically significant. However, the authors went on to attribute the greater success of the construction group as being almost entirely due to the non-selection of q. In the two tasks combined, the stimulus q was (correctly) omitted 19 times in the construction group, in the evaluation group, only 9 times; however, again in both tasks, both groups (incorrectly) omitted not-q 12 times. There was no difference in performance between the two tasks, that is, between the selection-task and the transfer task. "Knowledge is generalised to the extent that it has been gained" (ibid, p. 65).

The failure of half of the subjects to select not-q for examination was taken as confirming evidence for the theore-

tical model described above. Wason and Shapiro write:

> Its relevance [that of not-q] is assumed to depend on the arousal of a conflict between p-and-q (as reinforcing) and (not-q)-and-p (as falsifying). The conditions for this conflict occur if p had been selected and not-q omitted.

However, it is important to bear in mind that when the model says (box 10) "Card should be turned over; remove from LIST", the card is not physically turned over or removed from anything - the subject is not allowed to turn the cards at this stage. There is, of course, no reason why the subjects could not reverse this decision when conflict arises (though the model is plausible in that few actually do so). Now, since in the experiments described above, "...these conditions would be unlikely to have occurred because successive instances were constructed, or evaluated, independently of each other before the selection task is performed" (ibid), one would hypothesize (on this model) that the elimination of conflict results in the failure to select not-q. However, in both groups, half the subjects did select not-q, which is not explained by the model.

Following this relative failure to facilitate the gain of insight, Wason and Shapiro then turned to the question of whether "natural" experience would be more successful. The thematic material they used represented journeys taken on four days of the week, to Manchester and Leeds, by car or by train. Each card used had the name of one of the cities on one side, and a means of transport on the other. Subjects were presented the rule,

"Every time I go to Manchester I travel by car".

If destination and means of transport to that destination were matched on each card, then subjects should have chosen "Manchester" (p) and "train" (not-q) to test the truth or falsity of the rule. Ten out of sixteen subjects in the "thematic" group did choose correctly, while only two of a group using abstract materials chose correctly. The probability of this result being random was reported as being .004 under the one-tailed Fisher-Yates exact test.

The authors attributed the relative success of this form of the test to one or more of these causes: Firstly, the simple use of concrete rather than abstract material might have helped. As is well known, concrete materials are better remembered, and also inhibit fallacious inferences in syllogistic reasoning. In support of the latter, the authors cite Wilkins (1929)[10]. Secondly, the relation between the different terms ('travelling') is described as "concrete", aiding clear reasoning. (The present author is not clear on why "travelling from London to Leeds" is more concrete than "on the other side of the card", and the matter is not elucidated further in the paper under discussion, except to suggest testing subjects with rules such as "Every time I go to K I travel by 3"). Thirdly, the thematic material is said to form "a coherent, unified whole", leading subjects "to distribute their attention equally on

---

[10] This should be 1928, as is indicated by the title of the Wilkins paper. Both the text citation and the bibliographic reference are in error in the Wason and Shapiro paper.

its components, i.e. the four cards" (ibid, p. 69). Further, "With thematic material it is gratuitous to talk about combinatorial analysis: the activation of stored knowledge spontaneously generates 'real' alternatives" (p. 70). However, one is left with the feeling that familiarity with this kind of material provides the key, and that possibly one can not say much more beyond this. In a sense, the problem ceases to be interesting. Finally, reference is made briefly to Wason (1969b) in which the suggestion is made that "...it is not so much the logical structure which makes the abstract problem difficult, as the structure which the subject imposes on the problem" (Wason and Shapiro, 1970, p. 70). In Wason (1969b), it is stated that subjects who were presented with the solution first had no difficulty in justifying the solution. However, the actual experimental conditions in this case have never been described, in contrast to Wason's usual meticulous presentation. It is therefore difficult to judge his and others' (Johnson-Laird, 1971; Wason and Johnson-Laird, 1972) comments on this matter[11].

Two further "successes" were achieved in creating conditions under which subjects could solve the problem: in the first, using 'abstract' material (geometrical figures), subjects were required to discover the rule from a set of true or false instances, before being given the selection task

---

[11] The 1969(b) paper is otherwise a brief review article, and is not discussed further here.

with the same materials (Legrenzi, 1971); in the second, a possible 'real-life' situation was set up, in which subjects imagined themselves to be mail-sorters, and worked with stimulus materials which were envelopes, with or without stamps of two denominations, and sealed or unsealed (Johnson-Laird, Legrenzi and Legrenzi, 1972).

The conclusions from the former appear at first sight curiously subtle, for it is difficult to imagine how subjects could derive the rule, yet not succeed in the selection task. Nevertheless, one cannot disagree with Legrenzi's conclusion that discovery has aided the acquisition of insight. The experimental design was as follows:

Subjects were given four cards on each half of which was a triangle or a circle, and the cards were completely uncovered. In the order, 'circle/triangle, circle/circle, triangle/circle, triangle/triangle', the first only is a negative exemplar of the rule, "If there is a circle on the left, then there is a circle on the right". Subjects were told the first case was a negative exemplar and the remaining three positive exemplars of a rule concerning the arrangement of the triangles and circles, but they were not told the rule. The 'experimental group' subjects had to find the rule as their first task. (In a second form of the experiment, a further twelve subjects were asked to find the rule, and express it in the form, "If there is ... on the left, then there is ... on the right"). These experimental group subjects were then given four half-masked cards, and

were asked to pick out those cards they would unmask to show the rule true or false. The 'control group' subjects were assigned to each form of the rule produced by the experimental group in equal numbers. For example, if four subjects formed the rule, "The rule is true if there is a circle on the right, or two triangles together", then four control group subjects were given the rule in this form in the selection task. (In the second form of the task, of course, the rule was the same for all subjects).

In the first form of the test, 26 out of 30 experimental subjects got the selection task right. In the second form (all subjects using the "if ... then ..." formulation), 10 out of 11 experimental subjects succeeded. In the control groups, the corresponding figures were 14 out of 30 and 2 out of 11.

Two points arise from Legrenzi's description. Firstly, in the diagram showing the half-masked cards, the order of the cards, left to right, is the same as in the rule-finding case (all cards uncovered), shown in the same diagram. (The 'masks' are lightly-shaded, so that the reader can see the covered elements). It is not stated that the order was randomised, or changed, when it came to the selection task, and if this was not in fact the case, the impression is given that Legrenzi was trying to have his subjects succeed; such successes are by this time impressive! Secondly, it is surprising that such a high proportion of the control-group subjects were so successful in the first form of the experi-

ment. It is as though some subtle influence were at work in this experiment which had not been present previously.

The influence is, in fact, staring us right in the face. This is not the same experiment as the previous 'selection tasks'! There are only two stimuli, circles and triangles, whereas in all the previous work, each of two kinds of stimulus has been at least binary, e.g. red circles/blue circles; red triangles/blue triangles; or, vowels/consonants with even/odd numbers. The present situation is more closely analogous to the 'simplified' experiment of Johnson-Laird and Wason (1970), described on pages 18 to 21 above. The subjects then were also relatively successful.

The above argument has been presented in this rather theatrical fashion - essentially setting up the straw man provided by Legrenzi's experiment, only to knock it down in the preceding paragraph - for two reasons. First, it serves to emphasize the subtlety of the experiment in question; Legrenzi has, presumably in all good faith, set up the experimental situation, tested over 80 subjects, and written and published the report (which one assumes must also have been revieved by the Journal referees or editors), without recognizing that he had changed the experimental situation. And indeed, though "staring us in the face", it took the present author some time to spot the source of the anomaly. Second, this work of Legrenzi's has been cited as showing that 'discovery' does indeed lead to a deeper appreciation

of the logical structure of the selection task: see Johnson-Laird, Legrenzi and Legrenzi (1971); Wason and Johnson-Laird (1972); van Duyne (1974).

However, it must be emphasized that Legrenzi's contention has not been disproved. It is only to be regretted that the present author did not notice the oddity in Legrenzi's paper earlier, for it would have been a simple matter to test his claim with a small group of subjects.

In the second "successful" experiment referred to above, Johnson-Laird, Legrenzi and Legrenzi used as the rule for the selection task the two forms:

> If a letter is sealed, then it has a 50 lire stamp on it.
> A letter is sealed only if it has a 5d stamp on it.

Two parallel rules were used with abstract symbols (the letters A and D and the numbers 3 and 5). The letters and numbers were written on envelopes. Finally, the order of presentation was counterbalanced using Williams squares.

The degree of success in the realistic condition was striking. Of 24 subjects, 17 got both problems correct, 5 got one correct, and two got neither correct. In the abstract case, 7 of the 24 got one problem correct, and the remaining 17 got none correct. ("Correct" here means, of course, selecting p and not-q only).

The authors suggest that it could be the realistic <u>relation</u> between the contingencies in the rule which leads to the superior performance:

> ...The individual is used to considering the con-

> nexion between such items as destinations and modes of transport, or postal rates and envelopes (Johnson-Laird et al, 1972, p. 400).

They further suggest that "it is a task for the future to conduct direct tests of this explanation of the results, and to determine the precise nature of a realistic connexion between events. It may be that a sense of reality is no more than a feeling of familiarity" (ibid).

The present author is left with a feeling of vagueness equal to that hanging over this passage. What sort of explanation is this? And yet one feels that Johnson-Laird and his colleagues are close to the problem, in that we must define <u>what we mean</u> by a "sense of reality" before we shall be able to come to grips with the problem, a problem which goes far beyond the bounds of the selection task. For "familiarity" may enable one to cope with any one of a wide variety of problems studied by cognitive psychologists and their subjects, and the very word may lead us (by association?) to notions of stimulus-response associations, response probabilities, or cognitive nets. But are we any closer to an "explanation"?

<u>Previous work on the rule-finding task</u>.

The rule-finding task, of which the essence has already been described above (page 5), has received much less attention than the selection task, probably because the amount of variation which can be introduced is very limited. The difficulties subjects have with the problem were first des-

cribed in Wason (1960). Briefly, the following are commonly observed: Subjects produce a hypothesis, or guess, then assume it to be true, (occasionally insisting that it could not be wrong); subjects offer only verifying examples, which could never disprove their rule; subjects change their language in an attempt to change their strategy, but in fact repeatedly offer the same hypotheses as to what the rule is. For example, after suggesting that the rule is "Start with a number, then add that number twice", after giving more examples of the same pattern, the rule is offered that "the second number is random, then the first is half that number, and the third is three times the first".

Some attempts were described to inhibit hasty decisions, either by giving subjects money and then taking some away each time an incorrect rule was announced, or by encouraging subjects to announce other "possible rules", that is, other than that the experimenter had in mind (as the subject thought) (Wason, 1968a). In a further experiment, subjects were given only one chance to announce their hypothesis. If wrong, they were not told they were wrong, but were asked, "If you were wrong, how would you find out?" (ibid). Only four out of sixteen subjects suggested changing their hypotheses or generating instances inconsistent with them. Finally, to counter the criticism that the task might have an "intelligence test" quality, so that test-wise subjects might assume that only the "most fitting" rule was correct, one of Wason's students used the class of "living

things" as his rule, and a "Siamese cat" as his first example (Wason and Johnson-Laird, 1972, p. 212). Subjects were to generate other examples, and were told each time whether their example was a member of the class or not. It is reported (ibid) that "Very similar effects were obtained" - similar to those arising in the number-problem.

CHAPTER II.

EXPERIMENTAL PROCEDURES.

In this chapter are described the attempted replications of some forms of the selection task, and the few modifications introduced. The particular questions to which each form is directed will be raised as it is described. The two forms of the rule-finding task ('numbers' and 'animals') were given in essentially Wason's form, unchanged, and the procedure for these is described at the end of the chapter.

Part 1 - The Selection Task.

Three variants of the selection task, primarily, were used. Some modifications used later in the experimental phase will be described at the very end of this section. Four groups of subjects were used in the main experiments, and a fifth was added later, as will be described below.

Subjects.

The four groups of subjects tested were as follows:
(i) A group of "P.D.A.D." students from the Faculty of Education at the University of Alberta. These "Professional Diploma after Degree" students were taking a one-year program in education, having obtained already at least a first degree in another Faculty. Several had higher degrees, and several had apparently gained their first degree(s) several years previously, and might thus be classed as "mature" students. (Subjects were not asked their ages). There were 44

P.D.A.D. students.

(ii) A group of B.Ed. students. These subjects were mostly in the third or fourth year of the Bachelor of Education program at the University of Alberta, and had much less variance in age than the first group, being mostly about twenty-two to twenty-five years old. One or two B.Ed. students in fact already had another degree. Such people were classed with the P.D.A.D. group. There were 26 B.Ed. students.

(iii) A group of 45 high school students from Grades 11 and 12. It should be mentioned that the school from which subjects were taken was probably rather atypical in regard to the intellectual ability of its students. It was one of the Edmonton Separate School Board's three bilingual schools, and as such gathered a high proportion of its students from the children of professional families, children whose parents were keen that their children should become, or remain, fluent in a second language. Many of these students were thus very capable. This did in fact become clear during the present study, and was also demonstrated by the grades obtained by the students in their last semester.

(iv) The fourth group was a group of people with computer-programming skills. Some of these were employees of the Division of Educational Research Services at the University of Alberta, and some were acquaintances of the present author.

The reasons for selecting these subjects were as fol-

lows:

As stated above, previous workers tested mainly psychology students, but made reference to subjects from other areas. From 1972 in particular, more students were taken from colleges of education. However, since Wason and Johnson-Laird state that a wide variety of people have difficulty with such tasks (Wason and Johnson-Laird, 1972, 170-171), without reporting any detailed work with other groups, it was decided to include subjects other than psychology and education students.

The B.Ed. students (group (ii) above) were included for comparison with the subjects with similar backgrounds of Wason and his co-workers. The P.D.A.D. students, having generally a much more varied background than the B.Ed. students, were included for comparison with the latter - in particular, it might be thought that the science graduates would show greater analytical capacity, and perhaps, therefore, greater success at both the selection task and the rule-finding task, (contrary to the opinion cited above). Further, differences in approach to the problem might be revealed between subjects educated in the different disciplines, other than education.

The difficulties subjects are reported to have in attaining insight into the selection task, and even in recognizing their errors when all the information is made available to them, suggest a very strong form of mental 'set'. Consideration of the possibility that one may become less

flexible or less prepared to admit one's errors after several years of university incarceration, led to the inclusion of the high-school group.

Finally, modern technology has given rise to the growth of a class of "mental technicians" - computer programmers - who might be thought to be more capable of approaching the selection task in a logical fashion. Their craft[12] requires the constant use of logical statements, and the ability to follow a sequence of 'decisions' through a possibly complex program. If "transfer" does occur, such subjects should find the selection task trivial. For this reason, the fourth group above was also included[13].

During the course of the experiment, one subject suddenly changed her tactics. When asked afterwards why she had done this, she said she had remembered something from a course she was taking in "Logic in Teaching". As a result of the discovery that such a course existed, the author decided on a further variant, with the intention of testing subjects before and after this course when it was next of-

---

[12] To counter the possible charge that the choice of words in this and the previous sentence is in some degree pretentious, it should be mentioned that this choice was not casual. It seems to me that the activities of a programmer are more comparable to those of a skilled craftsman with physical materials than to any other working activity. The skill can be taught, it has limited boundaries, and frequent practice is required to maintain one's skill; yet there is art in its application.

[13] Professor S. Hunka has suggested that one group that might be even more likely to succeed than programmers would be designers of the logical circuits used in computers - people working at the "ultimate" "true-false" level. Circumstances prevented the testing of this hypothesis.

fered, and also of comparing success in the selection task with that in the course. It was thus hoped to investigate to a slight extent the teachability of logical skills, and, if these could be learned, to examine the effect of the course in logic on the particular skills required in the selection task. (In fact, the hope was not fulfilled because of the lack of volunteers from the class for the individual testing required in the selection task. Nevertheless, one or two points of interest arose, and the procedures followed with this group will be described below). Subjects from this course constituted the fifth group referred to above.

The three forms of the selection task: materials and procedures.

Each subject in the B.Ed., P.D.A.D. and high school groups received three forms of the selection task. Subjects in the computer programmer group received the first and third forms described below. The three individual forms will be described first, and then the assignment of subjects to the three forms, and lastly, the method of attempting to elicit progressive insight from whichever of the three a subject received last.

(i) The abstract form of the task.

The 'abstract' form used was identical in structure to those used by earlier workers. The stimuli were letters and numbers, and were binary. In order to avoid subjects cons-

truing "the other side of the card" as that side facing downwards, the stimuli were written on the same side of each card, and one of them in each case was covered with opaque black paper. The stimuli were the letters 'L' and 'X' and the numbers '6' and '7', and subjects were told that each card in fact had a number and a letter on it. Pairing the hidden character, in parentheses, with the displayed character, the cards used had on them

L(6) X(6) 6(X) 7(L)

It will be seen that only the case 7(L) falsified the rule.

Subjects were also given on a card, typed in large letters, the rule,

Every card with an 'L' on one half
has a '6' on the other half.

This card was kept exposed with the four stimulus cards during the tests.

If a subject asked a question such as "Does that mean a '6' must have an 'L' with it?", he was told, "The rule means just what it says. That's for you to decide".

Subjects were then asked to pick out those cards they would like uncovered in order to be sure the rule was either true or false for the four cases in question. Additionally, the printed form of this instruction was placed in front of them, and left there. This read,

"Choose those cases, and only those cases, you would wish to reveal to be sure the rule is either

true or false, for these four cases.
Do not ask to reveal too many or too few
cases". [A]

Though the rule refers to every card, and would therefore be falsified if not every card obeyed the rule, several subjects questioned whether the rule could be true in one case, yet false in another. As one subject said, "If I find a '6' with the 'L', can I then assume that it's true for all the cards?" Another put it this way: "Can I trust you to be entirely honest or entirely dishonest?" Such subjects were told that the rule could be true in one case, yet false in another. (Some said afterwards that they now realized they need not have asked). Similarly, subjects who by their comments or 'thinking aloud' indicated that they assumed that the 'truth' of one card implied the truth of all were cautioned that this might not necessarily be the case.

It might be argued that by giving such information, some subjects received enlightenment beyond the others. However, it would probably have been very misleading not to give this information, and it amounted to being sure the subjects understood the 'rules of the game'. Moreover, it was observed in almost all cases cases that subjects apparently did not have any particular insight causing them to ask the question; it seemed to be asked with little thought, and often was immediately answered by the subject himself. Finally, before the subject's choices were recorded, he was

asked "whether he would be sure, after looking at his chosen cards, whether the rule was true or false for these four cases". Thus, all subjects had this opportunity to re-assess their understanding of the task.

(ii) The 'journey' form of the task.

As stated above, two forms of the selection task are reported to have been found easier than the abstract task. In one form, the subject is invited to consider making a journey , and considers the test sentence, "Every time I go to Manchester, I travel by train". A little over half of the subjects were reported to have had complete insight. In the other form of the task, described as the 'realistic situation', subjects were told to imagine themselves to be mail sorters, and considered the rule, "If a letter is sealed, then it has a 50 lire stamp on it". The stimulus materials were envelopes, two showing stamps of denominations 40 and 50 lire respectively, and two not, of which one was seen to be sealed and the other to be unsealed. A fifth envelope, address-side up, had no stamp. Subjects could not tell whether the envelopes whose address-sides they could see were sealed or unsealed. In this experiment, 22 out of 24 subjects were successful.

Now, it is clear that the 'travel' form of the task also involves some degree of realism. Though not physically handling objects which might belong to a familiar situation, as in the mail-sorting task, the subjects would doubtless

have travelled to many cities by various means of transport. The experimental situation, though not 'real' was 'realistic'. Yet Wason and Shapiro (op cit) attribute the success of subjects in their experiment to its thematic nature, rather than to its degree of reality. However, it is possible to "immerse oneself" in a story in the realm of fantasy, and in the present work an attempt was made to separate the 'thematic' effect and that of reality by using, with one group of subjects, a travel theme in the unreal situation of interplanetary travel.

A second variant was also introduced with this problem in an attempt to see whether offering a third alternative (in contrast to the binary stimuli of the abstract version) would facilitate the acquisition of insight, and in particular, would prevent the mental conversion by subjects of "p implies q" to "q implies p". The effect was tested in both earthbound and space-travel situations.

The different forms were as follows:

Subjects were presented with one of the following rules:

"Every time I go to Lethbridge, I travel by car".

"Every time I go to Mars, I travel by spaceship".

The sentence was typed in large type on a card, which was kept before the subject throughout this part of the experiment. (No subject received, at any time, both sentences).

Subjects also had put in front of them, on the exposed halves of four cards, the following:

Lethbridge Calgary by car by air,

(for the first rule),

and

Mars Venus spaceship beam mass-transporter,

(for the second rule).

They were told,

> I want you to imagine that you travel frequently to Lethbridge/(Mars) and Calgary/(Venus), and that two means of transport are available. On each of these cards, a destination is given on one half [pointing] with the means of transport to that destination on the other half [pointing again]. That is, every card has the name of a city/(planet) on one half, with the means you use to get there on the other half. We are only concerned with these two destinations and these two means of transport.

The same test-cards and rules were used in the form in which a third destination and means of transport were included, but in this case, subjects were told that three destinations (Lethbridge/Mars, Red Deer/Mercury, Calgary/Venus) were involved, and three means of transport (by car/spaceship, by rail/rocket-ship, by air/beam mass-transporter). That is, the 'theme' involved three destinations, and means of transport, but the stimulus and test materials only showed the same two of them as with the other subjects. Subjects were thus free to imagine the third elements on the hidden halves of the cards.

In both the above cases, the possible destinations and means of transport were also stated on a separate sheet, which was kept available during this part of the test. Subjects were then asked to choose which of the four cards they would like to uncover to be sure that the rule was

either true or false for the four cases presented. They were also given the typed instruction, [A] (above, p. 51).

(iii) The 'realistic' form of the task.

As was mentioned in the review above of the Legrenzi work with envelopes and stamps as stimuli, it seems likely that it is impossible to present this material without directionality being involved in the instructions. An attempt was therefore made to use other practical 'realistic' stimuli. Foodstuffs, and utensils which could be used with these were chosen, specifically, some nuts, an orange, a paring knife, and a nutcracker. Two rows of four bags each were prepared on two shelves, one row above the other. Two bags in each row were closed, two left open. The nuts were placed in one open bag on the top row, the orange in the other; the knife was put into one open bag on the bottom row, the nutcrackers in the other. All the bags were "filled out" with a roll of paper, to keep open the mouths of the unsealed bags, and to prevent subjects guessing what was in the closed bags from the contours. The function of these rolls of paper was explained to all subjects.

Subjects were told that all the bags on the top row contained foodstuffs, those on the bottom row, utensils. They were then presented the rule,

"Every bag with nuts in it has a bag
with a nutcracker below it",

both verbally, and typed on a card. They were then assured

that "below" did mean directly, or vertically below. They were then asked which of the four closed bags [pointing] they would like to look into, in order to be sure the rule was either true or false, for these four cases. They were also given the typed form of the instruction [A, above].

In each of the above three cases (abstract, thematic, and 'realistic'), after the subject had stated his choice, he was asked, " - And do you think after you have looked at [his choice] you will know for sure whether the rule is true or false for these four cases?" At this point, some subjects realized they had not grasped the rule fully, and made a second attempt. It was their final effort which was recorded (i.e. when they believed they had understood the instructions).

The examination of choices and the elicitation of insight.

After the subject had made his final choice, and without moving the last set of stimuli used, or revealing the hidden elements, the subject was asked in turn, "from the left", the reasons for each of his choices. His responses were recorded on tape.

The subsequent course of the interview was determined by the subject's responses, but it was found difficult to follow the same fixed methodology with all subjects. Subjects vacillated, changed their minds completely, and argued (occasionally) with the experimenter - despite his

constant attempt to maintain a passive, gently-enquiring demeanour.

The general intention in the ensuing dialogue was to confront subjects with progressively increasing conflict, as their choices permitted. In the following, p and q will be used to refer to the first- and second-mentioned referent in each of the test-sentences. If their selection included not-p, subjects were first asked what would be the meaning of a q associated with not-p. (In fact, of those who selected not-p, most did so because they had chosen "all four" stimuli. At this point, many had the understanding that p implies q also requires that q implies p. If they lost this 'understanding', many then jumped to choosing only one or two stimuli; thus, the dialogue being described here does not follow a unique, experimenter-controlled sequence, in which irrelevant items are eliminated one by one, and the relevant items which the subject had omitted - usually just the not-q case - finally being included, with the simultaneous acquisition of insight. On the contrary, far from being tidily linear, the sequence was often circular). If the subject did not at this point reject the need to examine the not-p case, he was asked about the case of the q, (though occasionally the not-q, depending on his responses). Finally, the subject was always asked what would have been the meaning of a p associated with the not-q. All of this dialogue was recorded on tape.

The assignment of subjects to the tests.

Subjects in the B.Ed. and P.D.A.D. groups each received the three problems in their basic forms, that is, the travel form involved the two cities and means of transport, and subjects were told these were the only elements involved. If we label the abstract, thematic and 'realistic' forms A, B, and C respectively, three orders of presentation were used, A-B-C, B-C-A, C-A-B, so that each form occurred once in the first, the second and the third position. Subjects within each of the above groups were assigned at random to each of the three orders of presentation.

The High School students received four variants of the 'thematic' test, formed from the two travel situations (realistic or fantastic), each with either two or three destinations being mentioned. The subjects were randomly assigned to one of these four forms. In addition, the assignment to the three different orders of presentation was random, as in the cases above, that is, to the orders A-B-C or B-C-A or C-A-B. Only two forms of the test were presented to the programmers, the 'realistic' form, and the abstract form.

The subjects from the 'Logic in Teaching' course, and their testing.

The course which the subjects in this group were taking was a Spring Session course - a very intensive course lasting only three weeks. It was hoped to present the selection task to a number of these students in the first two or

three days of the course, and to others in the last two or three days. However, presumably because of the compressed nature of the course, only five students volunteered for the individual testing (and only one of these after the course). Furthermore, because they had been promised that these individual tests would be kept short, in view of the acknowledged time pressure on them, only two forms of the selection task were given, and the rule-finding task was omitted. The data from these subjects could not be analyzed in isolation, and it was considered unwise to include them in the P.D.A.D. or B.Ed. groups, who received three tests. Their data on the selection task was therefore rejected, though it should be stated that their performance was in no way atypical.

However, students from this class were given two tests of reasoning, before and after the course, with an acceptable sample size. The tests were the syllogisms test from the Kit for Cognitive Factors of French, Ekstrom and Price (1962), and an adaptation of Valentine's "Reasoning Tests" (1962). Fifty-five subjects took the syllogisms test, before the course, and of these, thirty-five took both the pre-test and the post-test of the reasoning test. The tests were split into two halves, and are presented in Appendix IV. In the pre-test, subjects were assigned at random to one of the halves of each test. In the post-test, subjects wrote those tests they had not received in the pre-test.

Part 2 - The Rule-finding Task.

Two forms of the rule-finding task were referred to in the literature review in Chapter I, that using numbers, with "2-4-6" as the starting-point, and that using members of the animal kingdom, with "A Siamese cat" as the starting-point. The same two forms were used in the present work.

Subjects.

Subjects tested were those of the B.Ed. and P.D.A.D. programs, and those High School students who had time to go on to this task within the class-period available for testing.

Assignment of subjects to the forms of the task.

Subjects were assigned initially at random to the two tasks. However, many subjects found the 'animal' form much easier, and finished it quickly. In so doing, they apparently made less information available about their cognitive processes as they proposed successive hypotheses. Some of these subjects - depending on the available time - were then asked to take the numeric form of the test. The converse, however, did not apply. Very few subjects succeeded in the numeric task quickly, and therefore few had time available for the 'animal' task.

Method.

(i) The numeric form.

Subjects were told,

> I am going to give you an example of three numbers which conform to a certain rule. I would like you to give me, one at a time, other sets of three numbers, and I shall tell you immediately whether your set agrees with my rule. Your task is to find my rule. However, please do not tell me what you think the rule is until you are quite sure you know it.
>
> The first example is "2-4-6". [This was written on the response-sheet, exactly as here, with the dashes].
>
> Each time you write down an example, would you please also write your reason for suggesting that particular example.

Subjects were given a sheet of paper with a column on the left for their suggested triads, and one on the right for their reasons for suggesting each example. They were assured that "The rule is simple, and mathematical ability is not needed to solve the problem".

If subjects gave two wrong rules, they were reminded once more that they "should only suggest a rule when they when they were quite sure they knew it". At the same time, a card was placed in front of them, on which was typed:

The rule is simple.

There are no tricks of wording.

It was explained that "no tricks of wording" meant that if they had clearly grasped the essence of the rule, however it be phrased, that would be acceptable.

The rule used was 'Any set of three numbers that are in ascending order'. As subjects offered the successive examples, they were told, "Yes, that agrees with my rule",

or "No, that does not agree with my rule". It was also suggested to them that they put a check-mark or a cross against 'right' or 'wrong' examples, so that they had a record in front of them.

CHAPTER III

Results.

Results are reported for all groups together, since there was broad agreement across them all, though the counts of levels of insight are tabulated separately, and of course particular differences will be mentioned explicitly. Between-group comparisons will be made afterwards. The results of the rule-finding task will be reported separately, no group comparisons being made. At the time this study was first proposed, it was hoped to create a numerical scale for the rule-finding task in order to compare group performances - especially between the science and mathematics majors and students from the non-mathematical areas, and in order to compare performance in the two types of task. As will be explained, to do so would have distorted the results to an unacceptable degree.

It is emphasized that numerical results are not considered critical to this thesis. The values attached to the 'degrees of insight' are necessarily crude, and insight itself turned out to be an ephemeral quality for many subjects. Thus the descriptions of subjects' behaviour may be more important than the analyses.

In all of the following, Problem A is the abstract problem; Problem B is the travel problem; Problem C is the 'concrete' form, using nuts, nutcrackers, a knife and an orange.

In each case, 'p' is 'L', 'Lethbridge/Venus', or 'nuts';

'q' is the '6', 'car/spaceship', or 'nutcrackers';

'not-q' is the '7', 'by air/ion-rocket', or 'knife'.

'not-p' is 'X', 'Calgary/Mars' or 'orange'.

Problem orders were:

Order 1 - ABC; order 2 - BCA; order 3 - CAB.

In Tables 1 - 4 are presented the number of choices in each of six categories for the three problem-types. It must be emphasized, though, that there was no evidence of simple pattern across the choices; for example, many of the 14 in the P.D.A.D. group who chose 'p and q' for problem B did not choose 'p and q' for problem A. Since there are 216 possible paths (6^3) through the three problems, and many different ones were in fact taken, it was not found possible to provide a brief description of sequences of choices.

Except for the frequency of selection of the 'p-only' case, the results for the abstract problem are similar to those tabulated by Wason and Johnson-Laird (1970), assembled from four of their earlier experiments to give their total sample size of 128. Thus, in the present case, in the P.D.A.D. group, 22 out of 44 - 50% - chose 'p and q'; the Wason and Johnson-Laird figure is 46%; 13% of the students

represented in Table 1 chose 'p, q, and not-q', the comparable figure from Wason and Johnson-Laird being 8%. However, there is a marked discrepancy in the frequency of choice of the 'p-only' case, 8 out of 44, or 18%, in the present experiment, 33% in the Wason and Johnson-Laird table. It is thought very likely that the small number choosing the case 'p only' in this research arises from the great efforts made to ensure that subjects understood the instructions, and especially that they could if they wished choose more than one of the four cases, since the misapprehension that they could only select one case seemed to be fairly common. A similar result was found in all groups, and in all three forms of the selection task, and the point will not be mentioned again. Within each of the groups, tests for an order effect and for differences in performance on the three tests were carried out by the unusual means of a chi-squared test. Usually, this test is invalid if used with individuals being counted in more than one cell. However, such correlations between the cells bias the test <u>against</u> the null hypothesis. Thus, if we do not reject the null hypothesis under these conditions, we are erring, if at all, on the conservative side. If we were to <u>reject</u> the null hypothesis as a result of such a test, then interpretation would be difficult, but the question did not arise.

It should be noted that an analysis of variance would introduce a similar difficulty of uncertain interpretation, for in addition to having a limited range of scores (say, 1,

2, and 3 for no insight, partial insight and total insight respectively), the scores are violently skewed - in all groups - because of the predominance of the 'no insight' case.

The data of Tables 1 - 4 are presented in Tables 5 - 8 condensed into tables of level-of-insight by problem (A or B or C) and by order of presentation. Below each appear the chi-squared tests for the null hypotheses,

H_0 :The level of insight achieved is independent of problem type, and

H_0 :The level of insight achieved is independent of the order in which each problem is received.

In several cases, too many cells dropped below the commonly-accepted limit of five for the 'number of observations expected' (Siegel, 1956, p. 110), and in these cases partial insight and total insight frequencies were combined.

We note that there does seem to be a slight learning effect in the P.D.A.D. group, in that 18 of the 44 subjects (41%) have achieved partial or total insight by the third trial. Quite apart from the fact that this distribution has a better than 10% chance of random occurrence, the progress is hardly startling. The same group also has a slightly higher success on the travel problem (B) than on the abstract and concrete versions, but it is surprising that none of the groups in fact approached the level of success reported by Wason and Shapiro (1971). A contributory factor to this discrepancy may have been an oversight made by the

present author: it was noted too late that Wason and Shapiro had labelled each of the four test-cards with the name of a different day of the week. This was not done in the present experiment, although the fact that the journeys could be made at different times was emphasized. (In the Lethbridge/Calgary case, it was suggested that "You might go to Lethbridge on Mondays and Wednesdays and to Calgary on Tuesdays and Thursdays"). In any event, the labelling of the cards may somehow have made the London group's stimuli more independent of each other, so that alternatives could be more clearly perceived.

The failure of problem C to raise the level of insight is not so surprising; although using concrete and related materials, the situation was nevertheless artificial. In the kitchen, if the nutcrackers are not in their appointed place (in the drawer below the nuts) one naturally looks in the next drawer along. As suggested above, it is thought that the Johnson-Laird et al (1972) arrangement using envelopes and stamps is not comparable with, say, the abstract form of the selection task. Some directionality in the instructions and the associated familiarity with postal regulations cannot be avoided, leading subjects to the 'not-q' selection as well as the 'p'.

Finally, the number of subjects gaining insight on questioning is tabulated against 'initial selection' in Tables 9 - 12. We note in particular that it is not possible to test the claim of Johnson-Laird and Wason (1972) in

support of their 'revised' model, that "many more of the Ss who initially selected only p gain complete insight than do Ss who initially selected p and q" (ibid, p. 144). The reason is simply that very few subjects in the present experiment selected p-only.

Within the High School group, the effects were explored of including the possibility of a third destination and means of transport, and of using realistic and fantastic situations. The results from these tests were analysed by means of a two-way analysis of variance. The A-factor was 'number of destinations' - either two or three; the B-factor was 'situation' - earth-travel or space-travel. For each subject, no insight was scored as 1, partial insight as 2, total imsight as 3. The cell means and the analysis of variance table are presented in Table 13, which indicates that we do not reject (at alpha=.05 level) either of the null hypotheses

(a) H : the level of insight is not affected by the number of destinations mentioned in the problem, and

(b) H : the level of insight is not affected by the reality of the situation - 'realistic' or 'fantastic'.

(It was hoped to include students' grades as a covariate in the above analyses, but the grades passed to the writer were all so uniformly high, with so little variance, that the notion was abandoned).

Group comparisons.

Within the P.D.A.D. group, a comparison was made between the science and non-science graduates. The line of demarcation was necessarily crude, but broadly, those whose 'major' subject was in mathematics, physics, chemistry or geology were included in the 'science' group. Any subject who had a biological science major and a science minor was also included in this group. In the chi-squared test, (Table 15), partial and total insight had to be collapsed, since otherwise four of the six cells would have had less than five as the 'expected number of observations'. It is seen that in both abstract and concrete conditions the null hypothesis that the two groups do not differ in level of insight can be rejected at the .05 level of confidence.

Finally, all four major groups were compared on level of insight, and again it was necessary to collapse partial- and total-insight cells. The abstract problem appears in Table 16, the concrete problem in Table 17. In each case, the null hypothesis that the level of insight does not vary systematically with group membership cannot be rejected.

The most striking qualitative observation in this experiment was the difficulty subjects had in understanding the instructions. Firstly, subjects frequently and very promptly selected just one card, in some cases because, as they said, they thought they could only choose one, in others because they thought they could look at that one,

then go on to others if they wished. This, despite the fact that all subjects had before them all the time the written instruction, "Choose those cases and only those cases you would wish to reveal to be sure the rule is either true or false for these four cases. Do not ask to reveal too many or too few cases". Prior to this, subjects had also been told verbally to pick out those cards - "and it might be one, or two, or three, or all four" - they would want to look at, to be sure the rule was true or false for these four cases, but that they were not to uncover them yet. Often, these instructions were not sufficient. It was as if subjects had developed a "temporary deafness" while listening to them. Another paraphrase was then tried:

> After you have picked the card or cards you want to look at, you should be able to tell me, without looking at them, that you will know for sure for all four cases whether the rule is true or false after you have looked at them. But don't pick too many or too few.

This long-winded explanation usually satisfied the subject[14]

One difficulty arises, of course, from the ambiguity of the English pronoun, "which" - "Which of these cards would you want ..." could mean one or more. The French is not ambiguous, since the words are distinguished: "Laquelle de ces cartes..." and "Lesquelles ...". An attempt to disambiguate the sense by saying "Which one, or which ones of these cards ..." usually produced confusion.

In general, it was observed that those subjects who did

---

[14] This way of expressing the problem also seems to me to bring out the essential unnaturalness of the situation.

particularly 'badly' in this task - which means chiefly in the 'questioning' section, of course, since the general performance was poor to this point, very frequently showed themselves to be irrational in subsequent discussion, and not necessarily concerning the experimental task. Subjects whose responses seemed 'emotional' or 'intuitive' continued in this vein. Subjects who had apparently been unable to lever out the appropriate concept through a lack of vocabulary seemed to be struggling with the same deficit in casual conversation. (Note that many subjects had received, or were close to receiving a degree from this university). The rule-finding task may be mentioned in this context, too. Here, the innumeracy of subjects, as revealed by their inability to find expressions such as "even", "increment", "increasing", (- "You know, getting bigger"), "integer", "limit" (in the sense of a bound) was frankly astonishing to the present writer. <u>These difficulties were not uncommon</u>, and many sufferers again had a university degree, or were close to receiving one.

Turning now to the rule-finding task, again Wason's results (1960, 1968) were broadly confirmed. Most subjects enjoyed the 'game', only one thought the very general rule "unfair", and yet a large proportion were not efficient in their strategies. As stated above, no attempt at quantitative comparison will be made. How can one compare, for example, the protocols of a mathematically 'naive' subject who gives four or five positive, integer examples, and then

announces the correct rule, with a subject who carefully covers a wide range of possibilities, changing his hypotheses each time, but makes one mistake, or overlooks a counter-example?

Once again, the only certainty that did arise was that success was unpredictable. Some mathematics and science students produced 'poor' protocols, some students of the humanities 'near-perfect' protocols. However, the comment of one student, with a B.Sc. in Chemistry is pertinent here, and was echoed by two or three other subjects: she claimed that in three years of chemistry, she "had not been allowed to think, only to memorize and regurgitate".

Whether it is due to nature, or to a lack of University nurture, some evidence of inability to reason was provided by the results of the logical and syllogistic reasoning tests presented in Appendix IV. Since the nature of this material is tangential to this thesis, the results will not be analysed here, save to say that they confirmed the general impression above - many students were <u>very</u> weak, some obtaining a negative score on the reasoning tests, out of a possible total of 50. No patterns were discernible - for example, questions with negative elements were not apparently more difficult than positively-framed questions. No learning effect was apparent across the three weeks of the course, but this is not surprising, since the course material covered a wide range, and it was not a course in "logical thinking".

CHOICE	1st PROBLEM	2nd PROBLEM	3rd PROBLEM
p only	8	7	6
p and q	22	18	15
p,q, and not-q	6	4	10
p and not-q	2	6	8
all four	3	5	2
other	3	4	3

SELECTION TASK: NUMBER OF SELECTIONS IN EACH CATEGORY.
P.D.A.D. STUDENTS.

TABLE 1.

CHOICE	1st PROBLEM	2nd PROBLEM	3rd PROBLEM
p only	3	3	1
p and q	16	11	12
p,q, and not-q	3	5	5
p and not-q	3	3	2
all four	2	3	4
other	1	1	0

SELECTION TASK: NUMBER OF SELECTIONS IN EACH CATEGORY.

B.Ed. STUDENTS.

TABLE 2.

CHOICE	1st PROBLEM	2nd PROBLEM	3rd PROBLEM
p only	2	2	0
p and q	20	19	19
p,q, and not-q	4	5	7
p and not-q	6	6	8
all four	4	6	3
other	9	7	8

SELECTION TASK: NUMBER OF SELECTIONS IN EACH CATEGORY.
HIGH SCHOOL STUDENTS.

TABLE 3.

CHOICE	1st PROBLEM	2nd PROBLEM
p only	2	2
p and q	6	4
p,q, and not-q	3	2
p and not-q	2	4
all four	1	1
other	0	1

SELECTION TASK: NUMBER OF SELECTIONS IN EACH CATEGORY.
COMPUTER PERSONNEL.

TABLE 4.

LEVEL OF INSIGHT	PROBLEM			PROBLEM NUMBER IN ORDER GIVEN		
	A	B	C	1	2	3
None	35	27	34	36	34	26
Partial	4	8	8	6	5	9
Total	5	9	2	2	6	8

For problem-effect: chi-squared=7.41; df=4; probability=0.12

For order-effect: chi-squared=6.57; df=4; probability=0.16

NUMBER OF CASES AT EACH LEVEL OF INSIGHT, FOR EACH PROBLEM:
P.D.A.D. STUDENTS

TABLE 5.

LEVEL OF INSIGHT	PROBLEM A	PROBLEM B	PROBLEM C	PROBLEM NUMBER IN ORDER GIVEN 1	2	3
None	19	21	17	20	18	19
Partial	5	3	5	3	5	5
Total	2	2	4	3	3	2

For problem-effect: chi-squared=7.41; df=2; probability=0.46*

For order-effect: chi-squared=0.39; df=2; probability=0.82*

*Partial and total insight-values were summed.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT, FOR EACH PROBLEM:
B.Ed. STUDENTS.

TABLE 6.

LEVEL OF INSIGHT	PROBLEM A	PROBLEM B	PROBLEM C	PROBLEM NUMBER IN ORDER GIVEN 1	2	3
None	31	33	35	35	34	30
Partial	7	6	3	4	5	7
Total	7	6	7	6	6	8

For problem-effect: chi-squared=1.97; df=4; probability=0.74

For order-effect: chi-squared=1.70; df=4; probability=0.79

NUMBER OF CASES AT EACH LEVEL OF INSIGHT, FOR EACH PROBLEM: HIGH SCHOOL STUDENTS.

TABLE 7.

LEVEL OF INSIGHT	PROBLEM A	PROBLEM C	PROBLEM NUMBER IN ORDER GIVEN 1	2
None	10	8	9	9
Partial	1	3	2	2
Total	3	3	3	3

For problem-effect: chi-squared=0.15 df=1; probability=0.72*
for order-effect: chi-squared=0.0 df=1; probability=1.0*

* Partial and total values were summed. Yates' correction was applied to the first case.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT, FOR EACH PROBLEM: COMPUTER PERSONNEL.

TABLE 8.

CHOICE	NUMBER WITH EACH CHOICE	NUMBER GAINING INSIGHT
p and q	15	3
p,q, and not-q	10	5
p only	6	0
other	5	2

NUMBER OF SUBJECTS GAINING INSIGHT AFTER EACH OF FOUR CLASSES OF CHOICE, IN LAST PROBLEM ONLY: P.D.A.D. GROUP.

TABLE 9.

CHOICE	NUMBER WITH EACH CHOICE	NUMBER GAINING INSIGHT
p and q	12	3
p,q, and not-q	5	2
p only	1	1
other	4	1

NUMBER OF SUBJECTS GAINING INSIGHT AFTER EACH OF FOUR CLASSES OF CHOICE, IN LAST PROBLEM ONLY: B.Ed. GROUP.

TABLE 10.

CHOICE	NUMBER WITH EACH CHOICE	NUMBER GAINING INSIGHT
p and q	19	6
p,q, and not-q	8	3
p only	0	0
other	11	1

NUMBER OF SUBJECTS GAINING INSIGHT AFTER EACH OF FOUR CLASSES OF CHOICE, IN LAST PROBLEM ONLY: HIGH SCHOOL GROUP.

TABLE 11.

CHOICE	NUMBER WITH EACH CHOICE	NUMBER GAINING INSIGHT
p and q	5	2
p,q, and not-q	2	2
p only	2	1
other	2	0

NUMBER OF SUBJECTS GAINING INSIGHT AFTER EACH OF FOUR CLASSES OF CHOICE, IN LAST PROBLEM ONLY: COMPUTER PERSONNEL.

TABLE 12.

Cell means for high school travel problem.

	Earth travel	Space travel
Two destinations	1.64 N=14	1.27 N=11
Three destinations	1.56 N= 9	1.45 N=11

Analysis of variance table

SOURCE	SS.	DF	MS.	F RATIO	PROBABILITY
SA	.025	1	.025	.038	.86
SB	.610	1	.610	.95	.34
SAB	.200	1	.200	.31	.58
SE	.26	41	.643		

CELL MEANS AND ANALYSIS OF VARIANCE TABLE FOR HIGH SCHOOL TRAVEL PROBLEM, 'REALISTIC' and'FANTASTIC' SITUATIONS, AND TWO OR THREE DESTINATIONS MENTIONED IN THE TASK.

TABLE 13.

LEVEL OF INSIGHT	SCIENCE GROUP	NON-SCIENCE GROUP
None	8	27
Partial	3	1
Total	3	2

Chi-squared=6.33; df=1; probability=0.01*

*Partial and total values were summed; Yates' correction was applied.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT: THE ABSTRACT PROBLEM.
COMPARISON OF SCIENCE AND NON-SCIENCE GROUPS.

TABLE 14.

LEVEL OF INSIGHT	SCIENCE GROUP	NON-SCIENCE GROUP
None	8	26
Partial	4	4
Total	2	0

Chi-squared=4.74; df=1; probability=0.03*

*Partial and total values were summed; Yates' correction was applied.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT: THE CONCRETE PROBLEM.
COMPARISON OF SCIENCE AND NON-SCIENCE GROUPS.

TABLE 15.

LEVEL OF INSIGHT	PDAD	B.ED.	HIGH SCHOOL	COMPUTER PERSONNEL.
None	35	19	31	10
Partial	4	5	7	1
Total	5	2	7	3

Chi-squared=1.35; df=3; probability=0.72*

* Partial and total values were summed.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT: ABSTRACT PROBLEM. COMPARISON OF THE FOUR GROUPS.

TABLE 16.

LEVEL OF INSIGHT	PDAD	B.ED.	HIGH SCHOOL	COMPUTER PERSONNEL.
None	34	17	35	8
Partial	8	5	3	3
Total	2	4	7	3

Chi-squared=2.24; df=3; probability=0.52*

* Partial and total values were summed.

NUMBER OF CASES AT EACH LEVEL OF INSIGHT: CONCRETE PROBLEM. COMPARISON OF THE FOUR GROUPS.

TABLE 17.

CHAPTER IV

DISCUSSION

I will attempt in this chapter to address myself to the following points, though of course many of the topics are closely interrelated and frequent cross-referencing may be necessary.

(a) Were the experimental findings of Wason and his co-workers corroborated?

(b) The tasks are frequently said to be "apparently" or "superficially" easy. Is this really true? Especially in the case of the selection-task, precisely what does the task consist of?

(c) Is the problem important? This is tied to two questions:

(i) How does the problem relate to other problem-solving situations investigated by psychologists in recent years?

(ii) Are the conclusions of Wason and his colleagues justified concerning their suggestions that Piaget's theory of a formal operational stage may need modification, that the selection-task is - I put it crudely here only to be brief - illustrative of the need in empirical science to falsify one's hypotheses, and that the rule-finding task reflects fundamental processes accounting for the Structure of Scientific Revolutions as described by Kuhn, 1970? This last has never been explicitly claimed, of course, but something close to this is implied in the repeated references to

Kuhn (eg. the four references in the brief Chapter 16 on the rule-finding task of Wason and Johnson-Laird, 1972; subjects' strategies were variously "consistent with", "in conformity with", "corroborating", "confirming" various aspects of Kuhn's thesis).

(d) Is the problem soluble, in the sense of answering the question, "Why do people have such-and-such difficulties?"

(e) Finally, we may ask what, if any, are the educational implications of this work, and is further research in the area necessary or worth-while?

Confirmation of previous work.

With minor exceptions, noted below, all of the findings of earlier workers were in fact confirmed. Subjects do have difficulty with the tasks, and once they start 'floundering', their thought-processes do indeed become erratic. There is ample evidence for this in the few protocols in Appendices II and III, which are quite representative. They have not been selected in any way for their dramatic value.

The particular difficulties or curiosities in behaviour caused by the immersion of subjects in the maze of these logical tests is already clear from the presentation of the results, and is illustrated by the protocols. The reactions of subjects do indeed seem quite perverse at times. Thus, why should subjects not take the advice to 'put a check mark

or a cross against each of your answers, so that you have a record of your earlier findings' (in the rule-finding task). Many did not, and a few even continued without such a record when they were clearly in difficulty, and after the advice had been repeated, because of their difficulty. One finds it difficult to imagine that people behaving in this fashion could even cope with life, let alone succeed at university. Clearly the problems are difficult in a rather unusual sense.

The difficulty of the tasks.

Many of the papers cited above have made early reference to the essential simplicity of the task - for example: "...a problem which is structurally simple..." (Wason, 1971, p. 206); "...What looks like a fairly simple problem is, in fact, a difficult one" (Wason, 1969a, p. 471); "These tasks are structurally simple but deceptively difficult..." (Johnson-Laird and Wason, 1970, p. 134).

Now of course, from a logician's point of view, examining the task as a logical exercise, it could hardly be simpler. Only making the choice purely binary could simplify it further. It should be emphasized, furthermore, that I am not disputing such comments as those just quoted. The problem is, in a sense, simple, and subjects almost uniformly have great difficulty with it. However, as I shall suggest below, it may depend on how one defines one's terms - the subjects do not after all know the structure of the task; to

them, its structure is the whole problem, as they perceive it.

What possible impediments, then, are met by a subject when the problem is first posed? I have been able to think of the following - there may be others:

(1) Does "if p then q" mean "if q then p"?, as commonly occurs in everyday life, giving rise to the 'defective truth table'.

(2) Does "'L' on one half" mean "'L' on the exposed half"? - in itself a trivial decision, but very easily assumed on one's first glance at the cards. (This is what Wason and colleagues have described as "irreversibility", which characterizes subjects as not being in the stage of formal operations). This is a very small point, and possibly even invalid since so many do choose the '6', yet it could add some small increment to the initial confusion.

(3) Does the request "prove true or false" mean the rule will <u>always</u> be true, or <u>always</u> false? (As I have stated above, the question need not be asked, since the rule always states, "<u>Every</u> card (time, etc)...", and one counter-example would be sufficient to break the rule. Yet, at least in this formulation of the problem, the adjective "every" might lead subjects to think in all-or-none terms).

(4) The subject is not allowed to turn over any cards while making his choices. He must therefore bear in mind, or thinks he must bear in mind, some choices while considering others. ("If I had found that one true/(false), need I look

at this one...?") While clearly not a problem for someone who has grasped the "simple logic" of the task, for those who have not done so, the problem rapidly becomes taxing - especially on short-term memory.

(5) There is some confusion for many subjects as to whether the rule is true or false for each card in turn or for all four simultaneously. This is logically the same as (3), but subjects do not see it this way.

Of course, if a subject were able to consider these questions individually and separately, and answer them, the problems would not arise. But this seems not to be the case. Having watched many subjects struggle to understand the task, and then make their selections, my opinion is that many subjects get involved in the above difficulties simultaneously, and then, if I may use a fishy analogy, in struggling to free themselves, pull the net of confusion tighter and tighter. Of course, this is a purely subjective argument, but may perhaps add a little insight into the difficulty of the task. I repeat, however, that the impediments above are in all probability not individually insurmountable by an 'average' subject; it is the confounding of some or all of them which gives rise to the difficulties. Note also that one way to remove the problem, and simultaneously make it more realistic, would be to allow subjects to turn over the cards they wished to as they came to them. In this way, subjects might recognize the irrelevance of the 'q' case as soon as they turn it over, and having turned the p-case, and

found it to be true, say, (so that search for disproof might continue), might also consider further the not-q case. I have tried hard to conceive of any situation one might meet outside the psychological laboratory in which one would have to withhold decisions or the testing of hypotheses in the fashion of this experiment, but have been unable to find one.

I now raise again the points made in Chapter I (page 8 and page 32) in which two and five subjects, respectively, were reported to be unable to comply with the instructions, or were seriously confused, or misunderstood the problem. That is, despite the "structural simplicity" of the problem, and the fact that most (other) subjects experience little "felt difficulty" (in contrast with the classical problems of Gestalt psychology - see Wason, 1971, p. 206), some university students find the comprehension of the task itself beyond them. One feels there must indeed be something very artificial about such a task. (I would also add that Wason's judgement - quoted on pp. 6 and 7 above - concerning subjects who claimed that 'not-p' and 'not-q' "already broke the rule" seems a little hasty or harsh. The concept of something following a rule was indeed "inadequately conceived" <u>in this instance</u>. The subjects had simply mistakenly leaped to the conclusion that the rule was <u>only</u> concerned with 'L''s and '6''s. However, this is a dire warning to a pedagogue not to assume that what he said was what his students heard! Similar cases of "temporary deafness"

arose in the present investigation, as was mentioned in the previous chapter).

However, this is not to deny any of Wason's (and others') conclusions. In particular, much of the interest of this work does not derive only from the difficulty subjects experience, but from their curious and often inflexible rationalizations after they have made their choices.

The importance of the problem.

Of course, the question of the importance of a problem, perhaps in psychology more than in any other area of investigation, must be almost entirely subjective. There are two interrelated aspects concerning the importance of the tasks under discussion: the relationship between these experiments and other "problem-solving", "rule-finding", "information-processing" experiments; and the interference in other experiments and in real-life tasks of the incapacities subjects reveal in the present selection and rule-finding tasks. My own (subjective) judgement is that the work of Wason and his colleagues is no less, but probably no more important than that in other problem-solving areas, with which I believe it to be only very weakly linked, and that the importance of the revealed incapacities may be exaggerated. In particular, Piaget's theory has been questioned for some time now on the grounds that most cognitive performance seems to be determined by context and previous experience as much as 'cognitive level'. (See, for example,

Bryant, 1974; Brainerd, 1975). The work of the 'London group' thus reveals another area which demonstrates the non-universality of Piaget's theory of formal operations, and Wason and Johnson-Laird may well be correct in concluding that the logical skills acquired by physical and biological scientists "are, in fact, only elicited by familiar tasks, and not cognitive skills which can be applied to any problem whatever. In other words, they are really practical rules rather than formal operations" (Wason and Johnson-Laird, 1972, p. 190).

The matter of looking for links[15] between the tasks studied here and other areas of investigation in cognitive psychology raises questions akin to those very searchingly posed recently by Allen Newell (Newell, 1973). He had been asked to comment on papers presented at a symposium on "Visual Information Processing". He complimented the authors, repeatedly and emphatically, on the high quality of their experimental research, but found himself forced ("Half of me, he says, "is distressed") to ask the question, rhetorically, "If any one of the speakers continued his present

---

[15] Wason has explicitly admitted the weakness of connections with other areas in cognitive psychology:
"A chairman might be expected to make some important generalizations about his research area. But in our field the crucial issues are ill-defined because the area itself is ill-defined. Instead, let us consider a problem which I made up..."(Wason, 1969b, p. 281).
"Current research on problem-solving is too fragmentary and lacking in cohesion to justify either a survey, or even more, a synthesis. Each investigator seems to be struggling for just a little insight. So I shall use this as an excuse to concentrate on my own work and that of my associates..." (Wason, 1971, p. 206).

rate of production of high-quality papers to retirement, ...where would psychology be then?" Psychology, he says, investigates phenomena (like the selection task); someone finds a phenomenon, and a "flurry of papers" investigating it follows. Then some of these phenomena lead to "oppositions" such as central/peripheral, serial/parallel, perhaps, in our case, simple conditions where success is possible, and conditions where it is unlikely. And, says Newell,

> As I examine the fate of our oppositions, looking at those already in existence as a guide to how they fare and shape the course of science, it seems to me that clarity is never achieved. Matters simply become muddier and muddier as we go down through time. Thus, far from providing the rungs of a ladder by which psychology gradually climbs to clarity, this form of conceptual structure leads rather to an ever-increasing pile of issues, which we weary of or become diverted from, but never really settle (ibid, pp. 288 - 289).

He goes on to quote a recent example from the respected journal, "Science", concerning some process in short-term memory in which the authors provide "yet one more explanation". "Regardless of the exact merits of their case, ...it can be stated with confidence that their article does not settle the issue... [Their work] provides good evidence for the general proposition that psychological issues have difficulty even fading away" (ibid, p. 290). Newell had made the "fateful error" of trying (at the request of the convenor) to "put them [the papers] all together", but he writes, "...Not only could I not put them all together, I did not see how they themselves were putting

them all together".

In this same fashion, Newell and Simon develop "models" for "Human Problem Solving" (Newell and Simon, 1972) - in which their computer-program "models" play chess and solve cryptagrams and simple logical problems[16], Berlyne and Garner produce "information-processing" models (Berlyne, 1965; Garner, 1962), which, while interesting and ingenious are for the most part divorced from reality. In fact, perhaps one of the merits of the selection-task experiment is its demonstration of the total incompetence and illogicality of most people when presented with a 'purely logical' problem.

Quite apart from the subjects' confusions, however, the element so often left out in such models is that of 'total' context (which slippery concept I will not attempt to define here) plus previous history. Thus, measures of redundancy in the English language, however many texts are extracted from, at whatever depth, will fail to catch the splendid and ubiquitous nuances in the writing of Nabokov. How can this be, since he is writing English? It happens because the reader takes to Nabokov his own experiences, linguistic and otherwise, which are far more varied than any textual word-count can encompass. And, in part, because Nabokov's experiences are fuller, richer and more numerous than ours, he

---

[16] This is not intended to detract from their achievements. On the contrary, at least at my present superficial level of understanding, I am very impressed with their book. However, Newell and Simon's models are not explanatory in the sense that physical science models are.

writes books for us to read, rather than vice-verse[17].

Similarly, it is the case that Wason's early interest in both the rule-finding task and the selection-task arose from his interest in the difficulties people were found to have with negation (Wason and Johnson-Laird, 1972), and yet his own work later showed that 'it all depended' - yet again - 'on context'. (See "In real life all negatives are false", Wason, 1972). Consider, for example, an examinee asked - out of the blue, as it were, which is to say in a 'barren' or 'sparse' context - to paraphrase

> It is not altogether unfitting for a study on negation to begin with a negative statement (Klima, 1964).

(Which he did). One imagines the examinee might have a little difficulty extracting the sense of this.

Now consider the following little scenario. Imagine a patient has been under intensive observation for some time, suspected of suffering from Lassa fever. However, the possibility has arisen of some alternative diagnosis, and further tests are being conducted. Now, the medical impressario strides in, rubbing his hands, and with a wry smile says,

> Gentlemen, it is not altogther improbable that our patient does not have Lassa fever.

Immediately, and without pondering or 'translating', we understand a sentence containing three negative elements.

[There is an interesting sidelight here concerning some

---

[17] It is just possible that Nabokov might accept the pun squeezed in here; I hope the reader will too.

early investigations of the difficulty of comprehending, remembering or transforming negative, passive, and negative-passive forms of sentences. Some adherents of transformational generative grammar theory claimed that extra delays, for example, in recalling negative sentences, involved operating on a negative marker in addition to the base or kernel sentence (eg. Mehler, 1963). However, Goldman-Eisler and Cohen simply did a careful count in several language situations of the frequency of occurrence of negatives, passives, and negative-passives, and showed that simple habituation could account for the delays. Simple, active, affirmative, declarative sentences dominated by far all other sentence-types (Goldman-Eisler and Cohen, 1970). The point of immediate interest is this: why had this count not been done in all the years before? Unwillingness, or unawareness of the need to falsify, following Wason, seems a likely contender. Indeed, it is very likely that the number of false trails to follow and of issues 'muddied' by frequent and hasty publication would be drastically reduced if authors would simply do their utmost to follow Campbell and Fiske's (1959) advice on convergent and discriminant

validation[18]].

Turning now to the relationship of these experiments with Popper's theory of refutability, or Kuhn's theory of the growth of science through revolutions, I would claim at once either that Wason and Johnson-Laird have only an extremely superficial knowledge of these authors, (especially Popper), or that the references are brought in, perhaps rather casually, to add substance and possibly stature to their work. (I think the latter is more likely). However, the following discussion of Popper's and Kuhn's work is not only introduced as a foundation for my criticisms of Wason and Johnson-Laird; for, as I try to show, an understanding of the processes by which science grows also illuminates the difficulties of psychological research.

For clarity, I first quote three comments of Wason and Johnson-Laird (1972) relating to Kuhn:

> There was, however, little evidence for the use of the falsification strategy, and hence the main interest lies between the variation of hypotheses (which of course frequently leads to implicit falsification of former hypotheses), and the verification of hypotheses by piling up confirming evidence for them. This provides some corrobora-

---

[18] This may be too much to ask. Witness Meehl, who writes in a slightly sad vein, concerning the controlling of nuisance variables, "I have sometimes wondered whether it is only in the inexact sciences that rather simple methodological truths have to be noticed afresh after the passage of an 'academic generation' or two. Does this strange phenomenon occur also in physics and chemistry? In psychology, one is uncomfortably aware of the truth of Gide's remark, 'It has all been said before, but you must say it again, since nobody listens'" (Meehl, 1970, p.394, ftn).

> tion, in an artificial task, for Kuhn's (1962) view that beliefs, or hypotheses, are only abandoned (if at all) when more adequate alternatives become available (pp. 207 - 208) [A].
>
> On more than half the possible occasions, the hypothesis is not relinquished even when it is known to be wrong. Time is needed to find a new idea in a large number of cases - a point which is again in conformity with Kuhn's (1962) views (p. 210) [B].
>
> ....Five of the eleven subjects first of all generated instances to confirm a hypothesis (in contradiction to their instructions), and only then attempted to eliminate them. This confirms, yet again, Kuhn's (1962) argument that the scientist carries out research with reference to a pre-existing 'paradigm' (p. 212) [C].

Now, each of these statements is essentially foolish in its reference to Kuhn, though the reason may perhaps not be immediately clear. It lies, I think, in the subtle relationship between Popper's theory and Kuhn's theory or description of scientific revolution or progress, for the Popper theory <u>explains</u>, in one sense, Kuhn's formulation, and simultaneously clarifies why there is none but the most trivial relation between the isolated psychological experiments referred to above, and Kuhn's work. I will try to show that there is in fact a <u>necessary</u> connection between situations described by Kuhn and Popper's description of the progress of science, and indeed of all knowledge, since Popper's solution to the problem of induction - the process of logically unwarranted generalisation by which we all seem, willy-nilly, to learn about the world around us - is that we learn, in fact, from our <u>errors</u>: a classification (to take just one example of our mental organizing) becomes

crystallized, so to speak, at the point at which it breaks down[19]. He writes,

> Are we rationally justified in reasoning from instances or from counterinstances of which we have had experience to the truth or falsity of the corresponding laws, or to instances of which we have had no experience?
>
> The answer to the problem is: as implied by Hume, we certainly are not justified in reasoning from an instance to the truth of the corresponding law. But to this negative result, a second result, equally negative, may be added: we are justified in reasoning from a counterinstance to the falsity of the corresponding universal law (that is, of any law of which it is a counterinstance) (Popper, 1974, p. 1020; original emphasis).

Later, Popper writes on the "Psychological and Pragmatic Problems of Induction":

> It is, I think, hardly open to serious doubt that we are fitted with an immensely rich genetic endowment which, among other things, makes us most eager to generalise and to look out for regularities; and also, to apply the method of trial and error. Now I assert that all learning of new things is by the selective elimination of error rather than by instruction (Popper, 1974, p. 1024; original emphasis).

(On this same topic, see also: F.A. Hayek, The Primacy of the Abstract, 1972; and K. Lorenz, The Innate Bases of Learning, 1969).

I have quoted Popper at some length here to ensure that this point is clear: falsification is for Popper an essential element in the everyday acquisition of knowledge, as in the acquisition of understanding in science.

---

[19] "Seem ...to learn" is in fact the operative phrase here, for Popper insists that we do not ("neither animals nor men") "use any procedure like induction, or any argument based on the repetition of instances. The belief that we use induction is simply a mistake. It is a kind of optical illusion" (Popper, 1974, p. 1015).

I now come to the essential point. Popper is by no means a purist, a 'theoretical' philosopher. He states repeatedly that one should not give up one's theories lightly, that progress has often been made by the 'patching' of theories, by the addition of "auxiliary hypotheses" (at least to the point where deeper clarity or a greater unity is achieved by reformulating the theory - perhaps as a result of a "severe test"). The requirement of testability is paramount - but after all, strong theories will often withstand strong tests. Now it is argued - (for example, by Lakatos; see reference above, p. 3) - that refutation is impossible; that "some suitable adjustment in the background knowledge" will always account for troublesome data. <u>But in a well-established science, this is not true</u>. Popper and Lakatos between them provide an illustration:

Popper had used the discovery of Neptune as an example of a possible refutation, starting from deviations which were observed in Uranus' orbit. Lakatos had then created a "characteristic story" of an unobserved planet, the presence of which was postulated to be the cause of such a disturbance, but apparently too small to be observed by telescope; bigger telescopes are built; a dust-cloud still hides the planet; satellites are launched... However, Popper points out that "...almost all possible kinds of misbehaviour - that is, all except a set of measure zero - would <u>not</u> be explicable by postulating the existence of a planet p'. Thus, Lakatos' 'characteristic story' is in fact <u>an</u>

extremely exceptional case" (Popper, 1974, p. 1007; original emphasis).

To illustrate, Popper is saying, for example, that planets following rectangular orbits, or orbits not obeying Kepler's Laws, would refute Newton's gravitational theory, but that just such a perturbation as would be caused by an unobserved planet obeying Newton's law would be 'infinitely' unlikely under any other (unknown) law.

We are now close to the essential distinction between psychology and the mature sciences, and to the so-called 'Kuhnian' behaviour of Wason's subjects. Neither psychology nor most of its sub-branches has any overall structure - the nomological net is almost entirely absent. Thus, almost any 'theory' can be 'refuted' by just the sort of reformulation referred to by Newell in his example from "Science" (or, equally, can be 'patched' by adherents of a theory); hence the propagation ad nauseam of Newell's 'oppositions' and unresolved issues[20]. Putting it metaphorically, the whole structure of psychology is rather like an amorphous fluid; the 'theories' are scarcely-delimitable, weakly-structured bodies of almost the same density as the fluid, and as each other. Any theory can 'grow' or 'move' - the fluid yields, other theory-bodies are barely disturbed, and the overall change is negligible. The contrasting mature-science picture would be that of a strong crystal-structure in which a disturbing theory might be a new ion, too large for the

---

[20] See Koch, Psychology as a science, 1974.

crystal lattice, from which distortions grow, slowly at first, spreading through the crystal as information and research activity spread through the active ('thermally agitated') research community. (A 'scientific revolution' would be the sudden transmission through the structure of a dislocation - sometimes resulting in cleavage!)

We are now in a position to see how the points which Wason and Johnson-Laird make regarding Kuhn's theory do in fact fit into the scheme of things in scientific research. The first point, included in quotation A above, is a reflection of the faith which scientists have in the pre-existing structure (and, of course, Kuhn put the matter more strongly than Wason and Johnson-Laird suggest in this passage: theories are abandoned - even when they have little explanatory value with regard to some phenomenon - only when more adequate alternatives are available). The reason is clear: a good theory already has much empirical support, has explained many facts, and 'fits in' with the whole edifice of science. 'Patching' or extending such a theory is not dishonourable - on the contrary, it is sensible, within limits. And what is to be gained by abandoning such a theory, when no alternative is available (as mentioned in quotation A)? The classical example here is probably the Rayleigh-Jeans radiation law, founded on the elegant and successful classical theory of electro-magnetism. Even when Planck did finally propose his 'quantised' version, this was 'clearly' absurdly ad hoc - for at that time quantum theory was not

even hinted at by most empirical results. Certainly before the Planck radiation law was formulated, what was the alternative (for most 'reasonable' men!) to clinging to the Rayleigh-Jeans Law?. Such situations also include the point in B above. Of course time is needed to make changes in an active, complex enterprise such as scientific research - for even though a hypothesis may be "known to be wrong", perhaps, in the real world, it can be adjusted or added to, to explain new, uncomfortable data. But to draw a comparison between this and the time needed by subjects to change their hypotheses is far-fetched in the extreme.

Similarly with the case of research "with reference to a pre-existing paradigm" (in quotation C): if a scientist were not working within an existing paradigm, he would not be doing science, for his paradigm is the pre-existing structure[21]. When embarrassing facts emerge, and alternate paradigms appear, it would be possibly foolhardy, and certainly intellectually taxing to make the effort to absorb the new structure (which might be wrong anyway); the old one might turn out to be equally rewarding, and possibly successful.

The position of the subjects studying the rule-finding problem is analogous, but they are operating essentially in a vacuum - there is no surrounding 'structure' to provide a possibility of reasonable refutation. In contrast, a

---

[21] It is perfectly possible for a non-scientist to be working within a structure. The reader might beware of the 'not-p' case here!

science based on some coherent structure has reason to seek other explanations only when the current structure does not provide an explanation for some phenomenon. A second salient point is that nature provides the scientist with phenomena to challenge his theories. Asking experimental subjects to provide their own counter-examples has nothing in common with the practice of scientists, actual or supposed, beyond the elementary description of the two processes. The unavoidable implication of the Wason and Johnson-Laird references is that the way scientists operate, if Kuhn is correct, is to some degree a consequence of the impediments imposed on their research activities by their inadequate mental processes. While this may be the case, the connection is certainly not proven, nor necessary.

In summary, then, I have tried to show that refutation, as a methodological ideal, makes sense only in the framework of a well-established structure; the strength or effectiveness of the method is proportional to the strength of the nomological net within which it is put to use, and that the usual, pragmatic use of the method - including defence of theories against attempted refutations - is at least consonant with Kuhn's description of the growth of science, and may explain the latter theory to some extent[22]. I have tried to argue that Wason's experiments show that educated

---

[22] Note that, although it does not, so far as I am aware, affect the argument above, I have not suggested that Popper believes that Kuhn is correct in all respects. I have not read Popper on Kuhn.

subjects are not very good at solving logical problems, but show nothing more; the characteristic errors they make, such as failing to try and refute, and failing to vary their hypotheses can have no necessary connection with Popperian refutation or with Kuhn's theory[23]. People, scientists included, may also be pig-headed, proud, lazy, and so on, causing them to be as inefficient in their professional work as Wason's subjects were in these tests, and causing them to cling to unprofitable theories for reasons other than their logical fallibility. But there is, again, no necessary connection between Wason's work and the supposed "Kuhnian" behaviour of scientists.

There is a second bird we may try to kill with this same stone of refutation-in-a-structured-context, although we shall need the help of a second stone.

I have up to this point touched only briefly on the association between the selection-task and Piaget's theory of a formal-operational stage, and have not considered at all the matter of the solubility of problems such as "Why do people have such-and-such difficulties?" Now both these questions are clearly tied to the problem of explanation in

---

[23] The criticisms I have made to this point of the work of Wason and his colleagues may seem gratingly ungenerous. I would therefore like to emphasize that within the limited context they have set themselves, the few errors noted above notwithstanding, I regard their experimental work and analysis very highly. In particular, Wason and Johnson-Laird's "Psychology of Reasoning" (1972) seems to me a model of clarity and continuity. In Newell's phrase they have "put it all together" uncommonly well.

psychology. I have myself felt the exasperation hinted at by Newell in the extracts above, and feel it most acutely at the end of psychological research papers, when the authors make statements such as,

"How is a particular way of saying something appropriate to the content one wishes to convey?" (Oldfield and Marshall, 1968);

"Inconsistency, when it is recognized as such, evidently enables the subjects to appreciate logical structure, and to correct error. But exactly how it does this is still unclear" (Wason, 1969a).

"More attention should be paid to the conditions which result in the disappearance of insight after it has been gained .... A problem is usually solved when its logical structure is clear. What extraneous factors militate for, and against, insight into their structure?" (Johnson-Laird and Wason, 1970a)

So frequently it seems to me, in this context, questions asking "how?", such as "How do subjects....?", or, "What mechanisms determine....?" are simply paraphrases for "I don't know what I am talking about". In what sense is this true? The answer is to be found by a comparison with question-asking in everyday life, or in science, in which, when a question is asked, the <u>questioner</u> has a frame of reference or structure, within which he expects the answer to lie. Thus if we ask a chemist to explain the reactivity of the -SH radical in cysteine, we <u>expect</u> him to reply in terms

of bond-strengths, ion sizes, and so on; if we ask an engineer how a particular thermostat works, we expect an answer based on familiar mechanical or electrical principles; if a nine-year old asks why minus one times minus one is plus one, there is usually not much point in telling him about inverse and identity elements and the distributive law in the field of integers. In this fashion, although a questioner does not (usually) know the answer to his question, he 'knows' or expects (again, most commonly) to be able to fit the answer into his pre-existing scheme of things, his present mental structure of the world.

The same is true of a practicing scientist. He "asks questions" of nature, usually, but not always, by experimenting[24]. We see then, that either in the acquisition of knowledge and understanding during individual development, or in the growth of the bodies of knowledge we call "sciences", new knowledge is added to an already-existing structure, and we cannot stray too far (in some ill-defined sense) from this structure without the question becoming meaningless.

Assuming, then, that I am right with regard to the examples quoted above, the small point arises as to how it is possible to ask such questions if they do not in fact arise from an underlying structure? I believe the answer is

---

[24] Astronomy is perhaps the prototypical case of a purely observational science. Nevertheless, as the above-mentioned discovery of Neptune indicates, it is perfectly respectable as a science.

simply that the questions (or similar questions, if some or all of these are not good examples) are <u>not real questions</u>; they are essentially parodying the 'hard' sciences.

But there is a much more fundamental question. Is it not simply possible that psychology is still too young a science to have formed a structure firm enough to support 'strong theories' and severe critical tests? After all, the subject-matter is dauntingly complex. Considering the enormous resources in manpower and money poured into the the enterprise in the last thirty or forty years, this might seem unlikely, but we should admit the possibility.

I would like at this point to try to do two things: firstly, I will try to present an argument in favour of the last position - that psychology is yet young; all we need is time and effort - by presenting a case which might reasonably seem to be analogous, and which we now know to be a well established scientific discipline, that of historical linguistics[25]. After this, I shall try to present in a very cursory fashion some of the analysis of the theoretical biologist Elsasser concerning current difficulties in cell biology, which seem in their essence to be close to those of contemporary psychology. There is a link between the two situations, but I cannot make it clear at this point. By using this device of two analogies, I hope to clarify in

---

[25] I shall not concern myself with any 'theories' of phonemic or grammatical change with time, but merely with the establishment of the different language groups.

some degree what seems to be the essential weakness of 'scientific' psychology in a manner I could not achieve by discussing psychology in isolation.

Historical linguistics is a purely observational study. Furthermore it is a study of the most complex of forms, of language changing in time, because of various factors, and frequently in geographical area of usage. Consider now the following example which we find in an introductory text (Hjelmslev, 1970):

We take the words for "moon" and "month", in this order, except where indicated, and look at their equivalents in various languages:

Gothic: mena, menoths (unvoiced)

Irish: mi (month)

Latin: mensis (month)

Greek: mene, men

Lithuanian: menuo, menesis

O.C.S.: Mesets ("moon" and "month");

(The second 'e' has a nasal sound like French 'fin')

Old Indic: mah ("moon" and "month")

"Who can doubt", says Hjelmslev, "the common ancestry of languages?"

In examining such words, each element is treated in turn, and 'rules' are arrived at for relating elements and languages. In the above list, the initial "m" is of interest. If the 'm' were post-syllabic, or terminal in a word, it would be treated differently by, or within some of

these languages. Thus High German 'd' is functionally equivalent (say the linguists) to the Germanic post-syllabic unvoiced 'th', and the High German 't' is equated to Germanic post-syllabic 'd', or the Gothic 'd'. In this way the Gothic opposition, 'brothar' (with 'th' unvoiced) and 'fadar' takes the form in High German of 'Bruder' and 'Vater'. Notice that we have moved away from the unequivocal case of the 'm' in the 'moon' and 'month' list. Different sounds are now equated, and this in words in which all the remaining phonemes may be dissimilar. But matters get worse.

As Hjelmslev points out (p. 24) "... a single expression element in one language can correspond to two or more different expression elements in the related languages... [and]... As we can see, a distinction found in one language may be obliterated in another". Thus the distinction described above is lost in Icelandic, ('brothir', 'fathir', both voiced), Danish ('broder', 'fader'), and in English ('brother', 'father'), "... since, [present emphasis] these languages have in this case only one expression element corresponding to both Germanic post-syllabic 'th' and 'd'...".

> Thus, genetically related languages by no means need have the same number of expression elements, or the same system of expression elements (ibid).

But this seems horribly ad hoc, rather like Wason's excuse of his own explanation (above, p. 18). It is as though we had evidence of languages evolving from common stock, illustrated by rules governing correspondences bet-

ween expression elements in different languages (such as the initial 'm' in the list above); however, when the rules are found not to hold, we refer to the misfits as "exceptions", or create a language "sub-family". In this way, a theory could never be disproved. (We might remind ourselves here that the development of a language cannot be observed).

Of course we now know that this is not the case, but how must the situation have looked to scholars working at the foundations of historical linguistics during the last century?[26] I am not wishing to imply that any newly-observed relationship would have shaken the discipline to its core, but the merits of rival theories could have been very difficult to assess, and any one theory, at that time, impossible to refute. The similarity with the situation in modern psychology is striking. However, certainty has been achieved in historical linguistics, except, presumably, for details currently being worked on, because of the overwhelming mass of facts which fit into the established theories, not in isolation, but in the manner of a tightly constructed, complex jigsaw puzzle. Particular linguistic features such as those mentioned above do not define language groups or subgroups in isolation, but in clusters of many elements, the presences of many different elements reinforcing each other. Similarly, of course, the simultaneous absence of a group of

---

[26] The argument would of course be stronger here if I could cite examples of actual feuds. Regrettably, however, I am not a linguist, but it is clear that such problems must have arisen.

elements from a putative language-group also confirms that group's coherence. The structure, or 'nomological net' is now so strong that it can withstand 'exceptions', though of course the effort to explain them will continue.

And so, it might be argued, is it not possible that the jigsaw-puzzle of modern psychology will likewise begin to come together? I think the analysis of Elsasser, to which I now turn, indicates that this is unlikely. Elsasser concerns himself with the description of biological states (macro-description), and prediction of behaviour (such as morphogenesis), from a detailed knowledge of quantum or micro-states (Elsasser, 1966). It should be emphasised, though, that the following argument is more analogical than that above concerning linguistics. This arises, as will be seen, because Elsasser is concerned with <u>uncertainty</u> arising from the presence of the myriads of quantum states compatible with any one macro-description. The definition of the informational uncertainty in psychology analogous to the quantum states is difficult, but nevertheless, as I shall

try to make clear, the abstract problem is the same[27]. (It should perhaps be stated at once that Elsasser does not have a solution to the biological version of the problem, but rather an indication - I am not sure whether it would be valid to call his argument a <u>proof</u> - that a solution is intrinsically impossible in terms of the classical scientific approach of deduction from a relatively few general principles). Elsasser's argument is closely-reasoned and necessarily complex, and I hope I shall not create too much distortion in the following precis of his thesis.

He states his aim thus:

> One of the chief purposes of biological theory as conceived here is <u>to specify the limitations of physical prediction</u>. [Original emphasis]. We mean by physical prediction in this context such prediction as is based on the laws of physics applied to models involving homogeneous classes (Elsasser, 1966).

He then points out that <u>all physical measurements</u> have to be made on homogeneous classes of elements - for example, large numbers of electrons, or large numbers of similar

---

[27] It is perhaps fair to assure the reader that the following discussion of Elsasser's analysis does have close abstract-logical and analogical relationships with psychological explanation. Especially if unfamiliar with Elsasser, he might otherwise consider that the present writer is about to wander down totally irrelevant pathways. This belief, though possibly more comfortable, is not the case!

However, I am omitting one conclusion of Elsasser's which may be important. Following Niels Bohr, he claims that there may in fact be <u>absolute limits on knowledge</u> which we may expect to find associated with organisms having a degree of autonomy: the two notions - autonomy and restrictions on knowledge - are said to be inextricably related. I am not discussing this idea because I do not understand it, but if correct, it explains at a very fundamental level the problems of scientific psychology.

molecules. Because of the uncertainty principle, any attempts to make measurements on individual atoms, electrons, and so on, are doomed to failure[28]. Thus it is emphasized that "science deals with propositions about classes and not with individuals outside of their being representative of classes" (ibid, p. 90). Now according to Elsasser, the approach to biology in which organisms are characterized as open systems, with multiple and complex feedback loops, which notions have also been applied to psychology, albeit vaguely - (see Miller, Galanter and Pribram, 1960) - has not been fruitful, and he attributes the outstanding difficulties to the existence of radically inhomogeneous classes of elements. This renders inapplicable, in systems of extreme complexity, the method of sampling of homogeneous classes, and, as Elsasser points out (page 45), this property is not restricted to biological systems, being purely a set-theoretic problem. Elsasser defends the existence of such classes (at least in a rather 'tight' sense) on the grounds that in a system as large and complex as a cell, there are an immensely large number of possible quantum-states the cell could be in - exceeding tremendously in fact, the number of possible cells there could be in our universe. The same is of course true for, say, a laboratory sample of a crystal. However, because of the vastly complex structure and dynamics of biological specimens, properties cannot be averaged

---

[28] (See the comment on the testing of human subjects below).

out the way they can in most physical systems; the detailed specification of a biological system is necessary if one is to predict its future behaviour. [See for example his page 46; the statistical arguments occupy Elsasser's chapter 2 (ibid)].

Elsasser now proposes that the existence of these finite, inhomogeneous classes "is the most significant abstract property of organisms", and hypothesises that it would lead to "observable regularities which have no direct equivalent in homogeneous classes" (ibid, p. 48). However, he points out (p. 47) that a formal proof must still be supplied that such regularities can arise in a finite universe populated by inhomogeneous classes, even though the existence of such classes with these properties 'need not' lead to a conflict with classical physics (pp. 49-52).

What are the consequences of such a situation? Elsasser writes:

> We may find the situation to be so complex that we would need many examples to elucidate it in detail, and we may then find that we run out of numbers of the class considered before a question regarding some regularities of their mutual relationship has been decided... It appears that intrinsic logical complexity becomes scientifically significant as a property of finite inhomogeneous classes" (p. 90).

Now this situation is clearly paralleled in psychology. Each individual is brought up in a totally different environment from other individuals, and each is subject to a myriad different influences. Furthermore, we are able to adapt to a remarkable degree to our environment - so that,

for example, laboratory results are hardly generalizable beyond the laboratory; the subject adapts himself to the experimental situation,and then goes out and resumes his 'normal' behavior. The problem of the adaptability of the human subject has thus given rise recently to a move towards "contextualism" or "situationalism" (Weimer, 1973; Jenkins, 1974; and of course, in undeveloped form in Miller, Galanter and Pribram, 1960; a counter-attack in the area of personality 'theory' comes from Bowers, 1973, who argues that the 'person x situation' interaction is the thing to study).

I thus claim that there is at least a similarity, and possibly a formal isomorphism between the difficulties of experimental psychology and those exemplified in the last quotation from Elsasser, above. One of the results of, in essence, trying to produce, by statistical artifice, a 'homogeneous' set of subjects has been criticized by Meehl, who discusses the "controlling" of extraneous "nuisance" variables, and exposes the counter-factual (i.e. absurd) conditions which result (Meehl, 1970).

The above argument could be considered too pessimistic. Biology, even cell-biology, is still progressing, and even if Elsasser is in principle correct, it may be that the effect of a 'finite number of radically inhomogeneous classes' has not yet been felt. Francis Crick has expressed some minor disagreement with Elsasser:

> [At this point] we need to clear up a point about biological explanation. This is always of two types. In examining every biological system, one can always ask how it works; meaning how, from a

> knowledge of its parts, one can predict its behaviour. Alternatively, one can ask how the system got that way; in other words, how it evolved.
>
> Now these two explanations are in fact very different, for this reason. Most biological organisms work rather reliably. Moreover, many of them can be obtained in almost identical copies. Hence, one can usually make predictable observations on them, at least as well as on many purely physical systems, such as, for example, the eddies in a stream.
>
> On the other hand, it is unlikely that the evolution of one species into another has this character. This depends on rare events (such as mutations) and what may be chance factors in the environment...
>
> Consequently, there is real doubt whether the actual process of evolution is predictable. It may be history rather than science (Crick, 1966, pp. 14 - 15).

We note in particular that "most of them can be obtained in almost identical copies" is directly contrary to Elsasser's thesis, and thus it seems that Elsasser and Crick may have a different universe of discourse. In other words, Crick is concerned with processes at a very much more gross level than Elsasser. However, it is more likely that Crick did not at that time appreciate that Elsasser's specific concerns were with those aspects of biology leading to the phenomenon of autonomy[29]. Nevertheless, all this does not affect the applicability of Elsasser's ideas as a model, to psychological processes, and Crick's second point is relevant here. The process of evolution, he says, may not be predictable, because 'rare' or 'chance' events may be the causal agents of change - but this is precisely the class with which Elsasser is concerned. In his terms, any parti-

---

[29] This is made clearer in Elsasser (1970), which of course was not available to Crick.

cular quantum state (occupied by, say, a cell) occurs extremely rarely, because there is an immense number[30] of other states the cell could be in.

Now if we can agree that the number of 'events' that can occur either to an isolated human subject, or to a group of subjects is similarly "immense", then psychology, like evolution, becomes history rather than science[31]. That is, we may have some idea how the psychological subject got to a particular state, once we know he is there, but have no means of predicting (in a scientific, as opposed to an actuarial sense) where he will actually go. We might also note in passing that there is a reflection of this problem brought to a clear focus in Chomsky's (1959) review of Skinner's "Verbal Behavior", when Chomsky notes that "reinforcers" (events increasing the rate of emission of operants) can never be identified until after the operants have been emitted. This occurs because the number of possible influences on verbal behavior is very large indeed.

Notice, now, the similarity and the difference between the points just made, and the situation in historical linguistics discussed above. Because of the immense number of variables involved, evolution may only make sense viewed as history. The development of language groups may also only make sense as history, but for the very different reason

---

[30] Elsasser loosely defines an "immense" number to be one whose logarithm is itself "very large".

[31] Gergen (1973) makes exactly this point with reference to social psychology. I am grateful to Charles Anderson for drawing this paper to my attention.

that the time-scale of interest is so long; once the groups are established, it is 'simply' a matter of teasing out the clues which yield the interrelationships. Although complex, the problem is not insuperable.

But now we can see in a better perspective yet another difficulty in analysing human behavior, which is that change itself happens at too fast a rate. Though not usually expressed this way, this is simply another way of looking at the reactivity of human subjects to investigation - as, for example, when a "pre-test" also turns out to be a "learning experience".

To summarize this section: links relating the work of Kuhn (1970) and Popper's theory of refutation with the responses typical of the subjects of Wason and his co-workers were examined, and found to be essentially empty, though plausible. The lack of progress, and of any sense of unity in psychology was discussed, and it was suggested that the overall structure of psychology was in general too weak to support refutation, and therefore that the <u>strengthening</u> of the structure was itself likely to be extremely difficult. There is of course some circularity in this argument, in that strong tests help to create stronger structures, and stronger structures permit more 'severe' tests - perhaps more general, as well as decisive. The possibility was discussed that this weakness is simply symptomatic of a very early stage of development. While it was not rejected, this explanation was considered unlikely, since there seems to

have been very little tightening of the structure - matters just become "muddier and muddier". Finally, a plausible explanation for this lack of structure was sought in an analogy with the work of the theoretical biologist Elsasser. This work suggests that conjecture and refutation may not be possible regarding certain aspects of biological systems, specifically those in which a degree of autonomy is displayed, in conjunction with rare states from an inhomogeneous set of states. It is conjectured that psychological subjects, who show a degree of autonomy, along with their infinitely-variable histories (a subset of the uncountable set of all possible histories) form just such a 'rare' sample. Hence, the suggestion is made, though very hesitantly, that psychological explanation in any scientific sense may be intrinsically impossible.

Suggestions for future research.

I turn first, very briefly, to the selection task.

Firstly, regarding the question of 'discovering' the rule facilitating insight, as indicated above, the Legrenzi (1971) experiment should be repeated with four abstract stimuli instead of two. This would establish (a) whether subjects can in fact discover the rule, and (b) whether such discovery also enables them to gain insight.

Secondly, Wason and others have repeatedly stated that 'it is as though subjects' reasoning were dominated by their perceptions'. In other words, reversibility is lost. Now,

Bryant (1974) has shown that young children can in fact cope well with many Piagetian tasks, such as transference, provided that the load on memory is minimal, and care is taken not to allow immediate perception to dominate the children's judgement. It would be interesting in the present case simply to cover the cards while subjects are thinking about the task.

Thirdly, it might be informative simply to let subjects turn the cards over as they wish (having arranged that only the 'not-q' case disprove the rule), and to ask them immediately afterwards about their choices. Though it is anticipated that similar errors would be made, it may be that the degree of fixation of the errors would be reduced.

As far as educational implications are concerned, the message is clear, demonstrated not so much by subjects' failing to solve the problem studied here, as by their curious rationalizations and their subsequent 'distorting' of instructions which were in fact quite clear. Though relatively uncommon, the latter is quite disturbing. Thus, educators themselves must be very aware of the limitations of 'logic' in explaining actions, solutions to problems, or in giving instructions. The indications are that Wason's generalizations are broadly correct, that in unfamiliar situations people are not logical.

The abstract problems which have dominated most of this chapter have troubled me ever since I first started studying

psychology. It seems to me of paramount importance to attempt to answer the question "What questions, or what kinds of questions, may we reasonably ask of psychology?" Unfortunately, as it turns out, "Only those which in principle can be refuted" is too simplistic an answer. If the analysis attempted above is correct, the structure which permits refutation is still lacking. Thus, it is implicit in the analysis put forward in this thesis that the pursuit of 'basic research' in educational psychology (which, of course, now includes much work in science teaching, mathematics teaching, the teaching and learning of reading, and so on), is essentially a misuse of valuable resources[32]. This is why this thesis is doubtless atypical in the meagre space devoted here to "suggestions for future research". To take a specific example, it is a fact that most mathematics students find the method of 'proof by induction' very difficult to grasp. Yet the rationale of the method is simple to state[33], and it is not the algebra of the proofs which give rise to the difficulties. (Indeed, students commonly prove a result correctly, see the result in front of them, yet still doubt their own proof!) Now if we ask why students have difficulty, beyond a simple lack of familiarity with this type of proof, what does the question mean? Until we

---

[32] Ebel made precisely the same point nearly 10 years ago. See especially Ebel (1967) p. 82.

[33] Assume the relevant expression true for some particular n, say k; show that under this assumption, it holds for k+1; then show the expression true for some particular n, and hence for all n. (n is integral).

can answer this and similar questions without beating about the bush and groping for impressive phrases, we are not practicing science.

But we must try to approach a solution to the problem of what questions we may ask, for otherwise the whole psychological enterprise consists of researchers shut away in their isolated boxes, endlessly chasing their tales, at enormous public expense[34].

Kafka smiles in his grave.

---

[34] Andreski (1974, pp 26-27) reflects on the futility of efforts in the social sciences, and contrasts their efficacy with that apparent in the enormous progress made in the physical and medical sciences. Professor Andreski, who is no radical upstart, being Chairman of the Department of Sociology at the University of Reading, England, is particularly cutting about the inefficacy of educational research in North America, especially in the United States.

Bibliography.

Andreski, S. Social sciences as sorcery. London: Penguin Books, 1974.

Apostel, L. The Relation between negation in linguistics, logic, and psychology. A provisional conclusion. Logique et Analyse, 1972, pp. 333 - 401.

Bartley, W.W. Theories of demarcation between science and metaphysics, in I. Lakatos & A.E. Musgrave, (eds.), Problems in the philosophy of science, North Holland Publishing Co., 1968, pp. 40-119.

Berlyne, D. Structure and direction in thinking. New York: John Wiley. 1965.

Beth, E.W. & Piaget, J. Mathematical epistemology and psychology. (Translator, W. Mays). Dordrecht, Holland: D. Reidel Publishing Co., 1966.

Bowers, K.S. Situationalism in psychology: An analysis and a critique. Psychological Review, 1973, 80, 307-336.

Bryant, P. Perception and understanding in young children. London: Methuen and Co., 1974.

Burt, C. The development of reasoning in school children. Journal of Experimental Pedagogy, 1919, 5, 68 - 77, 121 - 127.

Burt, C. Mental and scholastic tests. London: P.S.King & Son Ltd., 1921.

Campbell, D.T. and Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix.

Psychological Bulletin, 1959, 56, pp. 81 - 105.

Clark, H.H. & Begun, J.S. The Use of syntax in understanding sentences. British Journal of Psychology, 1960, 59, 219-229.

Crick, F. Of molecules and men. Seattle: Washington University Press, 1966.

Ebel, R.L. Some limitations of basic research in education. Phi Delta Kappan, 1967, 49, 81-84.

Elsasser, W.M. Atom and organism: A new approach to theoretical biology. Princeton, New Jersey: Princeton University Press, 1966.

Elsasser, W.M. The role of individuality in biological theory. in C.H. Waddington (ed.) Towards a theoretical biology, vol. 3. Edinburgh University Press, 1970.

Evans, J.St.B.T. Interpretation and matching bias in a reasoning task. Quarterly Journal of Experimental Psychology, 1972, 24, 193-199. (a)

Evans, J.St.B.T. On the problems of interpreting reasoning data: logical and psychological approaches. Cognition, 1972, 1, 373-384. (b)

Evans, J.St.B.T. & Lynch, J.S. Matching bias in the selection task. British Journal of Psychology, 1973, 64, 391-397.

Flores d'Arcais, G.B. & Levelt, W.J.M. Advances in psycholinguistics. Amsterdam: North Holland Publishing Co., 1970.

French, J.W., Ekstrom, R.B., & Price, L.A., A Kit of tests

for cognitive factors. New Jersey, Educational Testing Service, 1963.

Garner, W.R. Uncertainty and structure as psychological concepts. New York: John Wiley, 1962.

Goldman-Eisler, F. & Cohen, M. Is N, P, or PN difficulty a valid criterion of transformational operations? Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 161-166.

Goodwin, R.Q. & Wason, P.C. Degrees of insight. British Journal of Psychology, 1972, 63(2),205 - 212.

Hayek, F.A. The primacy of the abstract. in A. Koestler and J.R. Smythies (eds.), Beyond reductionism. London: Hutchinson and Co., 1972.

James, W. The Principles of psychology. New York: Holt, 1890.

Jenkins, J.J. Remember that old theory of memory? Well forget it! American Psychologist, November, 1974, 785 - 795.

Johnson-Laird, P.N. Linguistic complexity and insight into a deductive problem, in Flores d'Arcais (1970), 334-343.

Johnson-Laird, P.N., Legrenzi, P. & Legrenzi, M.S. Reasoning and a sense of reality. British Journal of Psychology, 1972, 63(3), 395 - 400.

Johnson-Laird, P.N. & Wason, P.C. Insight into a logical relation. Quarterly Journal of Experimental Psychology, 1970, 22, 49-61. (a)

Johnson-Laird, P.N. & Wason, P.C. A theoretical analysis of insight into a reasoning task. Cognitive Psychology, 1970, 1, 134 - 148. (b)

Klima, E.S. Negation in English, in J.J. Fodor and J.J. Katz(eds), The structure of language. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Koch, S. Psychology as a science. in S.C. Brown (ed.) Philosophy of psychology. London: Macmillan, 1974.

Kuhn, T.S. The structure of scientific revolutions. University of Chicago Press, 1970.

Lakatos, I. Falsification and the methodology of scientific research programmes, in I. Lakatos and A. Musgrave (eds), Criticism and the growth of knowledge, Cambridge University Press, 1970, pp. 91-195.

Legrenzi, P. Relations between language and reasoning about deductive rules, in Flores d'Arcais, (1970), 322-333.

Legrenzi, P. Discovery as a means to understanding. Quarterly Journal of Experimental Psychology, 1971, 23, 417 - 422.

Lorenz, K.Z. The innate bases of learning. in K.H. Pribram (ed.) On the biology of learning. New York: Harcourt, Brace and World, Inc., 1969.

Lunzer, E.A., Harrison, C. & Davey, M. The Four-card problem and the generality of formal reasoning. Quarterly Journal of Experimental Psychology, 1972, 24, 326-339.

Medawar, P.B. Is the scientific paper a fraud? in The Listener, September. 12th, 1963.

Medawar, P.B. The Art of the soluble. London: Methuen & Co., Ltd., 1967.

Medawar, P.B. Induction and intuition in scientific thought. Philadelphia: American Philosophical Society, 1969.

Meehl, P.E. Nuisance variables and the ex post facto design. in M. Radner and S. Winokur (eds.), Minnesota studies in the philosophy of science, Vol. 4. Minneapolis: University of Minnesota Press, 1970.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behaviour, 1963, 346 - 351.

Miller, G.A. The psychology of communication. New York: Basic Books, 1967.

Miller, G.A., Galanter, E., Pribram, K. Plans and the structure of behaviour. New York: Holt, 1960.

Newell, A. You can't play twenty questions with nature and win. In W.G. Chase (ed), Visual information processing. New York: Academic Press, 1973.

Oldfield, R.C., and Marshall, J.C. Language: selected readings. London: Penguin Books, 1968.

Park, C. The Soviet theory of set - An evaluation. Unpublished M.Ed. Thesis, University of Alberta, 1972.

Piaget, J. Une forme verbale de la comparaison chez l'enfant. Archives de Psychologie, 1921, 18, 141 - 172.

Platt, J.R. Strong inference. Science, 1964, 146, 347-353.

Popper, K. Conjectures and refutations. London: Routledge & Kegan Paul, (4th edn), 1972.

Popper, K.R. Replies to my critics. in P.A. Schilpp, The philosophy of Karl Popper. La Salle, Illinois: Open Court Publishing Co., 1974.

Siegel, S. Non-parametric statistics for the behavioural sciences. McGraw Hill, 1956.

Schilpp, P.A. (ed.) The Philosophy of Karl Popper. La Salle, Illinois: Open Court Publishing Co., 1974.

Smalley, N.S. Evaluating a rule against possible instances. British Journal of Psychology, 1974, 65, 293-304.

Valentine, C.W. Reasoning tests. London: Oliver & Boyd, 1962.

Van Duyne, P.C. Realism and linguistic complexity in reasoning. British Journal of Psychology, 1974, 65, 59-67.

Wason, P.C. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 1960, 12, p. 129-140.

Wason, P.C. 'On the failure to eliminate hypotheses...' - A second look. in P.C. Wason & P.N. Johnson-Laird, Thinking and reasoning, Harmondsworth, Middlesex: Penguin Books, Ltd., 1968. (a)

Wason, P.C. Reasoning about a rule. Quarterly Journal of Experimental Psychology, 1968, 20, 273-281. (b)

Wason, P.C. Regression in reasoning? British Journal of Psychology, 1969, 60(4), 471 - 480. (a)

Wason, P.C. Structural simplicity and psychological com-

plexity: Some thoughts on a novel problem. Bulletin of the British Psychological Society, 1969, 22, 281-284 (b).

Wason, P.C. Psychological aspects of inference, in Flores d'Arcais, (1970), 344-346.

Wason, P.C. Problem solving and reasoning. British Medical Bulletin, 1971, 27, 206-210.

Wason, P.C. In real life negatives are false. Logique et analyse, 1972, 17 - 38.

Wason, P.C. & Golding, E. The language of inconsistency, British Journal of Psychology, 1974, 65(4), 537 - 546.

Wason, P.C. & Johnson-Laird, P.N. Proving a disjunctive rule. Quarterly Journal of Experimental Psychology, 1969, 21, 14-20.

Wason, P.C. & Johnson-Laird, P.N. A conflict between selecting and evaluating information in an inferential task. British Journal of Psychology, 1970, 61, 509 - 515.

Wason, P.C. & Johnson-Laird, P.N. The Psychology of reasoning. London: B.T. Batsford, Ltd, 1972.

Wason, P.C. & Shapiro, D. Natural and contrived experience in a reasoning problem. Quarterly Journal of Experimental Psychology, 1971, 23, 63 - 71.

Weimer, W.B. Psycholinguistics and Plato's paradox of the Meno. American Psychologist, January, 1973, 15 - 33.

Wilkins, M.C. The effect of changed material on the ability to do formal syllogistic reasoning. Archives de

Psychologie, 1928, *16*, 1-83 (No. 102).

Woodworth, R.S. *Dynamic psychology*. New York: Columbia University Press, 1918.

Woodworth, R.S. & Sells, S.B. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 1935, *18*, 451 - 460.

APPENDIX I

Elements of the Propositional Calculus.

In this Appendix, only the very simplest terminology and rules are presented, sufficient for the understanding of their use in the present thesis.

Conditional sentences are of the form "If p then q",

eg. If he is an April Fool, then I am a March Hare.

Such statements are also said to be statements of material implication.

The predication p is the "antecedent"; the predication q is the "consequent".

The form of the sentence does not have to be conditional; for example, "All A are B" has the same truth table (see below) as "A implies B". However, in ordinary discourse, the forms are normally used differently, though not necessarily, of course:

All dogs are mammals.

If it is a dog, then it is a mammal.

The "If ... then" format is also used to denote causality in everyday conversation:

If she has otosclerosis, then she will go deaf.

However, though the class-inclusion form is clumsy, it is not forbidden grammatically:

All [persons who have otosclerosis] are [persons who will go deaf].

Truth Tables.

A single predication either is so, or is not so, eg. "p" or "not-p".

Two such predications and their negations can therefore be conjoined ("and")[1] in four ways: p and q, not-p and q, p and not-q, not-p and not-q.

A truth-table defines the truth or falsity of such sets of elements associated by "and" or "or", etc. Of course, 'higher-order' tables can be constructed in which one or both of p and q are themselves composed of more than one predication.

For the case of p implies q, (or 'either q or not-p'), the table is:

p	q	p implies q
T	T	T
T	F	F
F	T	T
F	F	T

Truth table for 'p implies q'

Figure 5

In particular, we note that only the conjunction of p with

[1] Because of problems of reproduction, no special symbols are used for "and", "not", "or", etc.

not-q can prove the implication to be false. However, the frequent occurrence of cases of class-inclusion in everyday speech, in which, in addition to "all p being q", many q are, in fact, p, seems to have given rise to the common acceptance of the equivalence of "if p then q" and "if q then p". This gives rise to the defective truth-table referred to in the text:

p	q	p implies q and q implies p
T	T	T
T	F	F
F	T	F
F	F	T

Truth table for 'p implies q and q implies p'

Figure 6

From the classical truth-table, only two forms of inference are valid:

p implies q; p, therefore q ('modus ponens'),

and p implies q; not-q, therefore not-p ('modus tollens' or 'contrapositive').

The other two forms of inference, both fallacious, have likewise been categorised; they are:

p implies q; not-p, therefore not-q ('denial of the ante-

cedent')

and p implies q; q therefore p ('affirmation of the consequent').

It is the last error, the affirmation of the consequent, that gives rise to the defective, but commonly-'used' truth-table[1].

---

[1] This is not intended to imply that subjects consciously or deliberately make this choice, but simply that they behave as though they were using this table.

APPENDIX II

Protocols from the Selection Task.

Subject 1.

This subject (a female mathematics major) is included for contrast with the following. Her performance is not typical. The subjects are first of all questioned about each of their four choices in turn (choose to look at, or reject). In the abstract case, S had chosen, correctly, to look at only the cards showing an 'L' and a '7':

"Why did you not choose the case with the '6' showing?"
"Well, if it had an 'L' on it, it would be true; if it had an 'X', it wouldn't matter, because it's only those with 'L' that have to have a '6' at the bottom."
"And why did you choose the one with the '7' showing?"
"If it has an 'L' at the top, then it's false". [Very precise. The easy recognition of the need to prove the rule false is most unusual].
"And why did you reject this card with the 'X' showing?"
"Because it doesn't matter. It is not 'L'".

Subject 2. Sociology and Anthropology major.
S chose to look at only the case with 'L' showing:
"Why did you not choose the '7'?"
"Well, I'd presume that since the ... it did not say that a

card with a '7' could not have an 'L' above it, but the card with an 'L' would have a '6' below, so to find if the rule is true, my obvious first choice would be to see if the 'L' and the '6' go together."

"And why did you not choose the '6'?"

"No particular reason. Just the 'L' happened to be right. (sic!). The '6' would have been just as good".

"And why didn't you choose the case with the 'X' showing?"

"Because the same holds true as with the '7'. The 'X' could have had a '6' on the bottom, which could not have told me anything, because it doesn't say it can't have. It is just that the 'L' and the '6' are connected". [N.B.]

"Alright. Now, suppose there were an 'L' under the '7'; would that make the rule true or false?"

"It would make it false, because 'Every card with an 'L' on one half has a '6' on the other half', and this one has a '7'".

"O.K. So if there could be an 'L' there, should we choose to look there or not?"

"No".

"You have said that if there is an 'L' under there [7] it would make the rule false, but yet you don't want to look at it?"

"No, I would rather find out directly and check with the 'L'".

"Alright; now suppose we look at this 'L' and find a '6' under there, which would mean the rule applies to this card,

nevertheless, if there is an 'L' above the '7', is the rule false?"

"Yes".

"So should we or should we not choose the '7'?".

"Oh, I see what you mean. Ah... Yes, it means that even if the 'L' and the '6' are connected, I can't trust you to follow through with the rule on the other cards, because it might not be so".

"O.K. So now would you like to choose the 'L' and the '7'?"

"I think I would choose the '6', and see whether it has an 'L' above it".

"Do you mean in addition to the 'L' and the '7', or in addition to the 'L' only?"

"Oh, I see. 'Every card...' [irrelevant remarks]; if <u>one</u> card does not indicate, then you have to see all four cards".

"Alright; now, let's take the case of the 'X' first of all... Now we are thinking about opening all four. If there is a '6' under this 'X', does that verify the rule, or does it falsify the rule?"

"It does neither: the rule does not say an 'X' cannot have a '6' underneath it, but an 'L' must have a '6'".

"O.K. Now, we were looking at the case of all four. Do we need to look at this one?"

"I don't think so", [said in an offhand, confident manner, having apparently forgotten that she had just before selected all four cards].

"Right, now we are back to those three, and in <u>this</u> case, with the '6' showing...?"

"It must be an 'L'".

"Suppose there is an 'X' above the '6'... "

"That doesn't invalidate the rule". [Prompt, and very clear, as though S has insight <u>case by case</u>].

"That does not invalidate the rule? Now, the question is whether the rule is true or false for all the cards. If this is an 'X' it does not invalidate it?..."

"That's right".

"... and if it is an 'L'?..."

"Then it validates it".

"Is there any point, then, in looking at this card?" [Experimenter is now clearly 'leading'].

"Well, if you are trying to find out whether the rule is true or not... [pause]... What's happened to my intuition this morning? I don't operate on the basis of logic". [Laughter]. S continues:

"You're right, it wouldn't matter what's in the top half. But <u>this</u> one is diagnostic, it has an 'L' on it, it must have a '6' on it."

"Alright let's take a look at the '7'... "

Subject 3. Male Commerce student. Showed complete insight in the transport problem, with correct reasoning clearly illustrated on the tape. However, the following shows his loss of bearings with the abstract problem.

S had chosen to look above the '6' only:

"I guess I'd like to look at this one ('6') and this ('L')... [Reads rule again]. Well, I don't want to look at this one ('X') ... [Great puzzlement - reads rule several times]. It's just the wording - it's quite similar to another one [problem] but not quite the same. I'd like to check this one too ('6')".

"O.K. How about the one with the 'X' exposed?"

"Well, if the 'X' had a '6', that would also contradict the rule. And I'll say that about this one ('7') too".

"So you'd like to look at all four of those cards?"

"I guess so".

"O.K. Now let me ask you... Suppose there were a '7' under the 'X'. Would that make the rule true or false, in this case?"

"If there was a '7'? - Well, it wouldn't say anything about the rule".

"And if there were a '6' there?"

"That would make the rule incorrect".

"Would you read the rule again please?"

[Reads aloud].

(E:) "So if we have an 'X' here and a '6' here [same card] what does that tell us about the rule?"

"It doesn't say anything, does it? - because we're only concerned with cards with 'L' on them".

"So you would like to change your mind, would you?"

[Affirmative mumble]. "Now, this case with a '7'. If it

had an 'X', how would it affect the rule?"

"It wouldn't say anything either".

- "And if it had an 'L'?"

"Then that would have an effect".

"Now if this ('6') had an 'X', would that make the rule true or false?"

"It wouldn't (sic), because we are only concerned with an 'L'".

"And if it had an 'L'?"

"It would make the rule true".

"So should we look at this case or not?"

[Reads rule again]. "No, it wouldn't" (sic).

It is difficult to appreciate that this subject has only minutes before shown complete insight on another problem, and explained his (correct) reasoning quite clearly.

Subject 4.

The responses of this subject show most clearly how inadequate a picture one would obtain by taking subjects' initial responses at face-value, perhaps simply coding them 'stimulus-response' fashion, instead of the using the (admittedly imperfect) 'clinical' approach adopted here.

The subject, apparently a mature and calm person, responding, initially at least, in a considered fashion, apparently shows complete insight in the second and third prob-

lems (and which are indeed scored as such in the present author's scoring-system), but yet shows under questioning that she is not in fact clear on the basis for her choices, and vacillates greatly in the subsequent dialogue:

[S has chosen to look at the bag below that with nuts exposed, and the one above the exposed knife - the 'complete insight' case]:

E: Why did you choose the case above the knife?

S: If there's nuts in that one, then your rule is false automatically, because there is no nutcracker in this bag.

E: And the case below the nuts?...

S: If there's a nutcracker, again in that one, then your rule is true, because there's nuts in the top one and a nutcracker in the bottom one.

E: And you chose not to look at the case below the orange...?

S: Well, there's no nuts in the orange (sic) so it would be false ...-ah, well, it wouldn't give you any criterion to go on, because there's no nuts in that bag.

E: And above the nutcracker? You chose not to look in that bag...

S: Well, for the first two cases [indicates the knife and the nuts pairs] - if you can see that, then this one obviously, if the rule is true, that bag has to contain nuts.

E: - but you remember you did ask if it can be true in one case and false in another [in the first experiment, with the letters and numbers; the experimenter inadvertently gives a

hint here of the parallelism between the cases, but this degree of similarity was either already clear to the subject, or she did not notice].

S: Oh yes! That's right. (Mumbles). Then that one could contain nuts, couldn't it? Yeah. O.K. [Light laugh].

E: You'd like to choose that case as well?

S: Yes. One would want that case, because if there are nuts in that bag, then your rule is true for this case...

E: And if there's an orange there?... [Neutral tone]

S: Then it's false.

E: Alright. [Pause]. Let's look at this one a little bit further. If ther's an orange in <u>this</u> bag, above the nutcrackers, then the rule is false?

S: Yes.

E: Could you read the rule aloud, please?

S: [Reads. Repeats sotto voce]. So that means that that bag would have to have nuts in it, if the rule is true.

E: It does? [Pause]. Let me ask you this question: what does the rule say about bags with oranges in them?

S: Nothing. [Very prompt]. Oh, it <u>could</u> have an orange in it then, 'cos it doesn't say that necessarily ... for an orange... there could be no nutcracker, or there has to be a nutcracker.

E: So this one could have an orange in it?

S: H'mm. [Affirmative].

E: And the rule would be, for this case...?

S: True.

E: Let me just ask you one more question: if that rule said, "Every bag with nutcrackers in it has a bag with nuts above it", would that mean the same thing?

S: [Repeats rule]. Yes.

E: That means the same thing?

S: Yes.

E: Alright, now here's a bag with nutcrackers in it. If that means the same thing, we have just said that a bag with nutcrackers in it must have nuts above.

S: So that bag <u>couldn't</u> have an orange then. [Volunteered, without prompting]. (Sighs). It would have to have... Oh, boy [Repeats rule. Pause, 45seconds].

It doesn't mean the same, because there is no mention of an inverse relationship in that rule... like, ... the inverse relationship that you gave me would not relate (sic) really, because there's no switching-back relation [? - word not clear on tape]. If the nutcracker and the nuts were directly related, it would have to say that "Every bag with nuts in it has a nutcracker below, or vice versa".

S. is now directed back to the letters and numbers case, her first problem:

E: Now, once again, from the left, you chose to look at the 'L' and the '6' and the 'X'. Would you like to revise your judgement? There's the rule. [Presents rule, and reads it aloud].

S: No.

E: You think that's correct?

S: H'mm. (Affirmative mumble).

E: O.K. Now, let me ask you about the case of the 'X'. If there were a '6' below this 'X', would the rule be true or false, just for this case?

S: [Pause]. It would be false.

E: Could you explain why, please?

S: Well, it says that every card with an 'L' on one half has a '6' on the other half, so for this ['X'] to have a '6' on it would prove that rule false. No. [Mumble. ..."Hold on"? Reads again.]. Well, the '6' could be here anyway [- below 'X'].

E: It could?

S: Yeah.

E: -so it would not prove the rule false?

S: No.

E: And if there were a '7' there, below the 'X'?

S: It wouldn't prove the rule false.

E: Do we then need to look at the case of the 'X'?

S: No.

E: Alright. So you'd like to look at just the 'L' and the '6'?

S: Yeah.

E: Alright. Now let's consider the case above the '7'. If there were an 'L' above the '7', would the rule be true or false for that case?

S: It would be false.

E: Should we then choose to look at the case above the '7'?

S: Yes.

E: So you would like to look at the 'L' and the '7' and the '6'?

S: [Pause, about 10 seconds]. Yes. [Uncertainly].

E: O.K.

S: Yes.

E: Right, now let's look at the case above the '6'. If there were an 'X' above the '6', would that make the rule true or false for this case?

S: It doesn't apply to the rule. 'Cos, ah... [Reads] ...no. If there was an 'X' on that card, then the rule would be false.

E: Alright. An 'X' with this '6' makes the rule false?

S: H'mm. (Affirmative).

E: O.K. Then let's consider the case with an 'X' and a '6' here [the case with the 'X' exposed]; wouldn't that make the rule false, then?

S: [Pause]. Yeah, it would make it false.

E: So, you think we should look at all four cards, do you? [Very gently].

S: Yeah. (Then, quietly, "I don't know". Mutters rule). Yes.

E: We should? O.K. Does this rule mean the same as "Every card with a '6' on one half has an 'L' on the other half"?

S: H'mm. (Affirmative).

E: It does? So if a card has an 'L' on one half, it's <u>got</u> to have a '6' on the other half, and if a card has a '6' on

one half, it must have an 'L' on the other half?

S: [Reads again]. Yes.

E: O.K. What does the rule say about "X's"?

S: [Promptly] Nothing. It says that an 'X' can either have a seven (sic) because, for an 'X' to have a '6' on it, would make that rule false, so the 'X' can only go with a '7' if that rule is to be true.

E: O.K. Let me now present you with two separate sentences - nothing to do with this - and ask you about the relationship between them:

A dog is an animal with four legs.

An animal with four legs is a dog.

S: Not necessarily.

E: They don't mean the same thing?

S: No.

E: A card with an 'L' on one half has a '6' on the other half.

A card with a '6' on one half has an 'L' on the other half.

S: [Without a pause]. No, it doesn't necessarily mean the same thing.

E: <u>May</u> a card with an 'X' on one half have a '6' with it?

S: Yep. [No hesitation].

E: You're sure?

S: No. [Laughter].

E: O.K. Thank you very much...

Subject 5.

This subject indicates on the travel problem the kind of responses one would 'expect' a reasonable person to make. Her performance in this and the rule-finding task was unusually good. S. was a P.D.A.D. student who had a degree in Fine Arts "almost entirely in Dance".

E: Why did you not choose the case of the car, N.?

S. Well, it wouldn't have mattered what city you went to by car.

E: And Calgary...?

S: We're not concerned with Calgary.

E: And why did you choose 'by air'?

S: Because Lethbridge there would be incongruent with the rule.

E: And you chose Lethbridge...?

S: Because that would show it true or false.

We now turned to the first test the subject had tried, the 'concrete' case, in which she had selected 'nuts' and 'nutcrackers' - the 'no insight' choice.

E: Would you like to consider again your first choice on this problem, N.? You chose nuts and crackers last time. Here is the rule... Do you think that was correct?

S: Yes. I'd choose the nuts and the knife.

E: Why is that?

S: Well, we only need look at the nuts in the foodstuff,

the orange is out.

E: And why not the crackers?

S: Well, the nuts are the predictor; it does not matter what is with the crackers. And if there's nuts with the knife, the rule is false.

S. was then asked to consider briefly her choice ('L', '7') in her second problem. She said "It's O.K. I'm only interested in disproving the rule". This was said quite without emphasis.

[This subject in subsequent conversation showed herself clear-thinking and level-headed - by no means universally true. Despite extensive prior experience - she had taught dance to young children, and to mentally retarded children, and had supervised for a year the running of a day-care centre, she had had difficulty in obtaining admission to the P.D.A.D. program because of her 'curious' academic background; it was apparently thought that her critical faculties might be insufficiently developed].

Subject 6.

This subject illustrates again the sometimes fleeting quality of insight in this problem. S. had chosen the 'L' and '6' in the abstract case:

E: 'R.' has just said she has just thought of something she hadn't thought of before. Would you like to tell me what that is, please, 'R'?

S: That these ['L' and '6'] are the only two cards that

could possibly have the 'L' and '6' together, because I see there's no 'L' on this card ['X'] and I see there's no '6' on this card ['7'], so by a process of elimination that left only the two cards.

E: You have told me why you chose 'L' and '6'; now let me ask you about the one with a '7' on it. If there were an 'L' on this card, would the rule be true or false?

S: (Promptly) The rule would be false.

E: Do you then have to look at this case?

(Pause). No. 'Cos it has to have the '6' to be true. [Reads rule again, aloud]. So you actually would have to see that one, 'cos if it does have an 'L' it shows that it is false. E: O.K. You are now changing your mind? You are saying that the 'L' doesn't necessarily have to have a '6' with it?

S: Yes.

E: O.K. Let's look at the case of the '6' here. If that card had an 'X' on it, would the rule be true or false for that one card?

S: It's irrelevant.

E: O.K. And if it has an 'L' on it?

S: Then it still doesn't prove anything; it just shows that so far, one part is right, but it doesn't prove that that rule has been taken into account... I couldn't know, with just that card, that the rule is true.

E: Would you then want to look at that card with the '6' on it?

S: H'mm [Affirmative]. Er, no, I guess it wouldn't matter, no, that one's irrelevant.

[The word "irrelevant", of this case, rarely comes naturally to subjects].

E: Could you explain why, please?

S: Because if it has an 'L' then it's true in this case, but not in all cases, but if it doesn't have an 'L', then it's irrelevant, because the first factor is that it have an 'L' on it.

This subject, apparently having gained complete insight, then turned to the foodstuffs case and promptly rejected the 'knife', the critical choice. Furthermore, she changed her mind twice as to whether she needed to look above the nutcrackers - the case analogous to the '6' in the above.

APPENDIX III

Protocols from the Rule-finding Task.

Subject 1.

This subject, whose protocol, though short, is not perfect, demonstrates some of the difficulties of scoring these responses.

2-4-6	
1-3-5	Differences of two.
4-6-8	Differences of two, origin unimportant.
3-6-9	Even spaced, not by two.
2-4-8	Not equal distances.
1-2-3	Lower to higher. [Note that here, S. says this, but is not testing it; he is probably anticipating:]
1-3-2	[No comment]
1-5-69	No other relation. i.e. same relation does not hold from 1 - 5 as from 5 - 69.

The subject at this point announced the correct rule, but had not tested negative numbers, rational, irrational and transcendental numbers. S. was a mathematics major (P.D.A.D.), and had no difficulty with the selection task. He paused for some time after '1-3-5' and then said, "Am I to do this in the minimum number of choices?"

Subject 2.

S., with a background in geology and physics, had great difficulty with the selection task. Here he shows a mixture of good and poor strategy, with the apparent need to 'confirm' his results.

2-4-6	
8-10-12	Even numbers with a difference of 2 between them.
14-16-18	As above.
50-52-54	As above.
1-3-5	Three numbers, n1, n2, n3, with n2-n1=2; n3-n2=2.
11-13-15	As above.
4-2-6	Given three numbers so when rearranged, (if needed) they would be of the form N2-N1=2; N3-N2=2.
6-4-2	[No comment]
5-3-1	Increment of two increasing.
2-6-10	[No comment]
1-10-19	[No comment]
19-10-1	[No comment]
2-4-8	Trying to eleminate [sic] possibility of having constant number between any two numbers in succession.
2-8-12	As above. Rule announced.

This subject demonstrates the common reluctance of many subjects to give reasons for their choices. We also notice

that his first negative example, 4-2-6, was 'accidental' in giving the clue to the ascending order.

Subject 3.

This subject found the 'Siamese cat' form of the task easy, yet was not especially atypical:

Siamese cat	
German shepard dog	
	Kind of animal [- but we notice that country is still included].
horse	animal.
gopher	Wild animal.
car	machine.
pencil	inanimate object - not animal. [S. said after, that she meant 'not mobile' also]

Subject 4.

This P.D.A.D. student had time for both forms of the test, and shows the difference in the relative difficulty of the two.

Siamese cat	
a chair	to eliminate animate objects.
a lion	to establish the category of the example.
a fish	As above.
a bird	As above
a desk	As above
A tree	As above. Correct rule announced.

We notice that S. was 'changing his hypotheses', but not implying this within his 'reasons'.

The numeric case:

2-4-6	
8-10-12	increasing the first number in the set by increments of 2.
7-9-11	No comment.
2-8-16	The numbers must be all even or all odd. Asked if he was saying this as 'the rule', S. said he was.
(-2),(-3),(-4)	No reason.
(-2), (-4), (-6)	Rule: The numbers must be positive numbers.
102-104-106	'No real reason'.
one-third, -half, -quarter	eliminate fractions. Rule: whole numbers above zero.
0-0-0	eliminate 0.
2-7-6	One, to establish the hiearchical [sic] nature of the sequence[? S. Probably means 'ascending'; but quite what the 'one' means is not clear].
2-8-6	No reason. Rule announced.

After learning that '0-0-0' was wrong, S. said, "I'm eliminating things wholesale here, and I'm not eliminating them properly". Thus, he recognizes the importance of eliminating elements, and <u>had used this strategy</u> with the

'animal' task.

Subject 5.

S. had a 'double major' in Fine Arts and Urban Sociology, and had performed well above average on the selection task. The task took her 13 minutes, and her strategy is almost optimal. The previous subject (numerical example above) took 35 minutes.

Siamese cat.	
Lion	Cat family
animals with fur.	As above. Experimenter asked for particular example, yielding the non-cat response:
bear with fur.	includes more than cat family.
horse	no reason given.
monkey	No reason.
Eagle - 2 legs.	To verify that it includes more than just those animals which could be found in a zoo. On learning correct, S. wrote in, "includes more than only 4-legged animals".
	Rule: Classification includes all animals which have at least four legs, and definitely those which have two. When asked, S. said she did specifically exclude snakes and crawling things, but

added, "Now I know what I forgot to ask".

snake — has no legs - it crawls.

fish — to check whether animals live only on land.

amoeba — to check size.

giraffe — to check size.

Rule announced. S. said that 'picture of an ape on two legs broke the four-legs association'.

APPENDIX IV

The Reasoning Tests

Presented to the 'Logic in Teaching' Class.

Syllogisms Test.

This is a test of the ability to tell the difference between good reasoning and bad reasoning in a very limited context.

Some people find syllogistic reasoning difficult, others find it easy. If you find it difficult, do the best you can; everyone can accomplish something. Please work carefully through the examples, but do not start the test until asked to do so.

Instructions and examples.

Each item consists of two statements that are followed by four conclusions. It will be your task to examine each pair of statements, and to decide which one of the four given conclusions is the correct one.

Here is an example:
No birds are insects.
All swallows are birds.

Therefore:

A. No swallows are insects.
B. Some birds are not swallows.
C. All birds are swallows.
D. No insects are birds.

Since insects include no birds, and birds include all swallows, conclusion A is correct. You would therefore circle the A against the correct conclusion.

Here is another example:

All loans are profitable.
Some loans are investments.

Therefore:

A. All profitable things are investments.
B. Some profitable things are loans.
C. Some investments are profitable.
D. Some investments are not profitable.

Investments include some loans, and ALL loans are profitable. Therefore conclusion C is correct.

NOTICE that a correct solution is derived from both statements, and from those statements only. A correct conclusion is not just a repetition of the contents of just one of the statements, nor is a correct conclusion based on information other than that supplied by the given statements.

Please do not turn over until you are asked to start.

You will be given ten minutes for this test.

1. No footballs are round.
   All handballs are round.

   Therefore:
A. No handballs are footballs.
B. Some balls are not round.
C. All handballs are balls.
D. Some round things are not footballs.

2. No salesmen are bashful.
   All hucksters are salesmen.

   Therefore:
A. Some salesmen are not hucksters.
B. No hucksters are bashful.
C. All hucksters are brash.
D. Some bashful individuals are salesmen.

3. All little girls are cute.
   Some little girls are fat.

   Therefore:
A. Some little girls are not fat.
B. Some cute individuals are not little girls.
C. No fat little girls are cute.
D. Some fat individuals are cute.

4. No educators are stubborn.
   All teachers are educators.

   Therefore:
A. No teachers are stubborn.
B. Some educators are teachers.
C. All teachers are reasonable.
D. Some stubborn individuals are educators.

5. Some novels are classics.
   All classics are recommended readings.

   Therefore:
A. All novels are recommended readings.
B. Some recommended readings are novels.
C. Some classics are not novels.
D. Some recommended readings are not classics.

Turn over.

6. All fire engines are red.
No hearses are red.

Therefore:
A. No fire engines are red hearses.
B. Some red things are not fire engines.
C. All hearses are black.
D. No hearses are fire engines.

7. All divers are swimmers.
Some divers are sailors.

Therefore:
A. Some sailors are swimmers.
B. Some swimmers are not divers.
C. All sailors are swimmers.
D. Some divers are not sailors.

8. No railroad engines are airplanes.
Some railroad engines are coal-burners.

Therefore:
A. No airplanes are coal-burners.
B. Some coal-burners are not airplanes.
C. Some railroad engines are not coal-burners.
D. No airplanes are railroad engines.

9. All citizens are voters.
No minors are voters.

Therefore:
A. Some citizens are minors.
B. All minors are non-voting citizens.
C. No minors are citizens.
D. Some voters are not minors.

10. All artists are creative.
Some scientists are not creative.

Therefore:
A. Some scientists are not artists.
B. No artists are scientists.
C. Some creative individuals are not artists.
D. Some scientists are creative.

Turn over.

11. No executives are timid.
All managers are executives.

Therefore:
A. All managers are timid.
B. No timid individuals are executives.
C. Some executives are managers.
D. No managers are timid.

12. Some chisels are dull.
All chisels are tools.

Therefore:
A. Some tools are dull chisels.
B. Some tools are chisels.
C. Some tools are dull.
D. Some chisels are not dull.

13. No territories are self-governing.
Some islands are territories.

Therefore:
A. Some territories are not islands.
B. Some islands are self-governing.
C. No self-governing areas are territories.
D. Some islands are not self-governing.

14. All primates are apes.
No apes are monkeys.

Therefore:
A. Some apes are primates.
B. No monkeys are primates.
C. Some primates are not monkeys.
D. Some monkeys are not apes.

15. All soldiers are men.
All sergeants are soldiers.

Therefore:
A. All sergeants are men.
B. Some soldiers are sergeants.
C. Some men are not sergeants.
D. Some sergeants are not men.

Turn over.

16. Some workers are productive individuals.
All productive individuals are assets.

Therefore:
A. Some productive individuals are not workers.
B. All assets are productive.
C. Some workers are not assets.
D. Some assets are workers.

17. Some engineers are designers.
All engineers are graduates.

Therefore:
A. Some designers are graduates.
B. Some engineers are not designers.
C. Some graduates are engineers.
D. All designers are graduates.

18. All mistakes are errors.
No solutions are errors.

Therefore:
A. Some errors are mistakes.
B. No solutions are mistakes.
C. Some errors are not solutions.
D. Some solutions are not mistakes.

19. No generals are youngsters.
Some soldiers are youngsters.

Therefore:
A. Some youngsters are not soldiers.
B. Some soldiers are not generals.
C. No soldiers are generals.
D. No generals are soldiers.

20. No businessmen are federal employees.
Some businessmen are gamblers.

Therefore:
A. Some businessmen are not gamblers.
B. No federal employees are businessmen.
C. Some federal employees are gamblers.
D. Some gamblers are not federal employees.

Reasoning Test.

FORM A

This booklet contains two sets of tests.
Part I has two tests, Part II has 6 tests.
You are strongly advised not to spend more than 13 - 15 minutes on Part I.

If you finish Part I before this time, carry on with Part II; you may have time to return to Part I later, if needed.
You will be given 35 minutes for the whole test.

The items should be attempted in the order given, as in each part the harder tests come towards the end. However, do not spend too much time on any one question. Pass on to the next item, and return to the previous one later, if you have time.

Thank you for your cooperation.

PART I

1. John Huggins was found shot dead in Church Street at 1.45p.m. He had two known enemies, Bill Frogger and Jack Toper. The doctor says it could not have been suicide, and that Huggins died between 12.30p.m. and 1p.m. the same day.

Frogger was seen running from Church Street by two reliable witnesses at 1.10p.m.; Toper was seen by two reliable witnesses two miles from Church Street at 1.10p.m.; and Frogger was seen by two other reliable witnesses in a tavern one and a half miles from Church Street at 12.30p.m. Neither Frogger nor Toper had any means of transport except their own legs. The maximum speed of running for each was 1 mile in 8 minutes.

Assuming that none but the two named could have committed the murder, underline what must be true in the following statements, cross out what must be false, and put a question mark in the box by those which may or may not be true.

Frogger (a) was the murderer.....__
(b) could have been the murderer.....__
(c) could not be the murderer.....__

Toper (a) was the murderer.....__
(b) could have been the murderer.....__
(c) could not be the murderer.....__

2. There are four towns, A, B, C, D. A is the same distance from B that B is from C, and C is half that distance from A.

D is the same distance from C that C is from A.

Is A nearer to C or to D, or the same distance from each?

Diagrams may be used.

PART II

Examine the following arguments and state whether they are sound or not. You must assume first that the given premisses (i.e. the statements underlined) are true. The problem is, in each case, this: granted that these assumptions are true, is the other statement necessarily true? If you think the argument is sound, underline "Yes" and cross out "No"; if the argument is unsound, cross out "Yes" and underline "No".

If you say that the argument is unsound, show which of the sentences (i to iii or iv) given below that argument, gives the best reason why the conclusion does not follow from the given premisses. Mark your selected reason with an X in the blanks provided.

You are advised first to decide for yourself whether the argument is sound or not, before examining the reasons given below it. If the argument is sound it is obviously useless to examine the reasons which follow it.

Remember that you must assume that the underlined statements are true.

NOTE: Marks will be deducted for wrong answers, so that mere guessing is penalized.

3. "All successful authors are industrious. John Smith is an industrious author. Therefore he is or will be a successful author".

(i) It is not true that all successful authors are very industrious.
(ii) The fact that all successful authors are industrious does not imply that all industrious authors are successful.
(iii) Some successful authors are both industrious and clever.

Yes. No. (i).....(ii).....(iii).....

4. "None but Whigs vote for Mr. B. All who vote for Mr. B are ten-pound householders. Therefore none but Whigs are ten-pound householders".

(i) Only Whigs vote for Mr. B, yet all Whigs need not do this, so that there may be Whigs who are not ten-pound householders.

(ii) All those who vote for Mr B. are both Whigs and ten-pound householders, yet there may be ten-pound householders who do not vote for him, and hence need not be Whigs.

(iii) Even if none but ten-pound householders vote for Mr B., that is not to say that some of the ten-pound householders do not vote for his opponent, and hence are not Whigs.

(iv) There may be voters who are not Whigs, yet who vote for Mr B. on personal rather than on political grounds, and these will also be ten-pound householders.

Yes. No. (i).....(ii).....(iii).....(iv).....

5. "All the students are either industrious or intelligent. Either industry or intelligence will ensure success in the examination. So all the students will pass the examination".

(i) This conclusion is incorrect, for a student may fail in the examination through misfortune; for instance, he may feel unwell when the examination takes place.

(ii) A student may be industrious or intelligent, yet may be disqualified in the examination for bad conduct, eg. copying from other candidates.

(iii) Either industry or intelligence alone is surely insufficient for a success. A combination of the two is needed.

(iv) If a student is neither industrious nor very intelligent he may pass if by good fortune he is asked questions bearing on the little knowledge he has.

Yes. No. (i).....(ii).....(iii).....(iv).....

6. "If all the accused were innocent, some at least would have been acquitted. We may infer then, that none were innocent, since none have been acquitted".

(i) The innocent are often condemned to suffer for the guilty. Condemnation is no proof of guilt.

(ii) The guilt of some may have placed all in a bad light, so that none would be acquitted.

(iii) If only some would have been acquitted in any case, then some were not acquitted when they ought to have been.

(iv) We are only told that some of the accused would be acquitted if all were innocent. The number innocent may have been less than all, and so insufficient to secure any acquittals.

Yes. No. (i).....(ii).....(iii).....(iv).....

7. "No soldiers should be brought into the field who are not well qualified to perform their part; none but veterans are well qualified to perform their part; therefore, none but veterans should be brought into the field".

(i) If only veterans were brought into the field, young soldiers would never have a chance to learn, and when the veterans died there would be no one to replace them. The conclusion is, therefore, unsound.

(ii) Soldiers could not become veterans without going into the field as recruits, so the whole argument is false.

(iii) It is a misstatement to say that none but veterans are well qualified to perform their part, since young soldiers make up for their lack of experience by their enthusiasm.

(iv) Veterans may not be well qualified to perform their part for they may be too old, in which case the conclusion is invalid.

Yes. No. (i)..... (ii)..... (iii)..... (iv).....

8. "In Tutland only Conservatives - and not all of them - are Protectionists (i.e. against free trade); only Liberals - and not all of them - are Home Rulers: but both parties (Conservatives and Liberals) contain supporters of Women's Franchise".

It may be assumed that:
1. No Liberal is a Conservative;
2. all who are not Protectionists are Free Traders;
3. all who do not support Home Rule are Unionists.

This is all that is known about the views of the Tutlanders.

Hence, with this information, it would be incorrect to conclude that only the Unionists are Protectionists.

(i) Unionists need not be Conservatives; they may be Liberals, so that although Conservatives are Protectionists, Unionists need not be.

(ii) In the given premisses, Unionists are found only in the Liberal Party which contains no Protectionists. Thus Unionists cannot be Protectionists.

(iii) Only the Liberals are Home Rulers. Therefore all the Conservatives must be Unionists. As only Conservatives are Protectionists, it follows that only Unionists are Protectionists.

(iv) The questions of Unionism and Protection are entirely independent of each other. Thus no conclusion can be drawn as to whether believers in Protection are also believers in Unionism.

Yes. It would be incorrect.
No. It would not be incorrect:
(i)..... (ii)..... (iii)..... (iv).....

Reasoning Test.

FORM B

This booklet contains two sets of tests.
Part I has two tests, Part II has 6 tests.
You are strongly advised not to spend more than 13 - 15 minutes on Part I.

If you finish Part I before this time, carry on with Part II; you may have time to return to Part I later, if needed. You will be given 35 minutes for the whole test.

The items should be attempted in the order given, as in each part the harder tests come towards the end. However, do not spend too much time on any one question. Pass on to the next item, and return to the previous one later, if you have time.

Thank you for your cooperation.

PART I.

1. At a dinner Mr A had soup, fish and cheese
Mr B had soup and fish but no cheese
Mr C had fish and cheese, but no soup.

Nothing else was eaten or drunk. Later Mr A and Mr C developed food poisoning. (We have no report yet about Mr B).

Assuming that the cause of the poisoning was in the dinner mentioned, underline those of the following statements which are certainly true, cross out those which must be false, and put a question mark by those which may be true or may not be true.

There was

(i) poison in the soup
(ii) poison in the cheese
(iii) no poison in the fish
(iv) poison in the soup and the fish
(v) no poison in the fish or the cheese
(vi) poison either in the fish or in the cheese
(vii) poison in the soup, the fish and the cheese
(viii) poison in the fish and the cheese.

2. A man left his money to his five sons as follows; find the scheme or principles on which he divided the money, the individual personalities not being considered.

To A, aged 35, with 2 children, and income of $400, he left $500
To B, aged 40, with 3 children, and income of $500, he left $700
To C, aged 45, with 1 child, and income of $400, he left $600
To D, aged 35, with 2 children, and income of $600, he left $500
To E, aged 30, with no children, and income of $300, he left $200

(Half marks will be awarded for a partial solution of this problem, i.e. if only one principle or rule is discovered. For the full solution, precise figures must be given).

PART II.

3. "Everyone is either well-informed of the facts or already convinced on the subject. No-one can be at the same time both already convinced on the subject and amenable to argument. Hence it follows that only those who are well-informed of the facts can be amenable to argument".

(i) This conclusion is the converse of the true one, for those who are well-informed of the facts will be sure of their ground and so will not be amenable to argument.

(ii) A man may be convinced on the subject, yet if a good argument is ably put to him, he may alter his opinion.

(iii) There is no reason why everyone should be either well-informed of the facts or already convinced on the subject; there may be people who have never heard of the subject at all.

(iv) The first premiss is not clear, for some may be both well-informed of the facts and already convinced on the subject, and according to the second premiss, these will not be amenable to argument.

Yes. No. (i)..... (ii)..... (iii)..... (iv).....

4. "If you argue on a subject which you do not understand, you will prove yourself a fool; for this is a mistake fools always make".

(i) The statement is not sufficient to define "a fool". Only one characteristic is given, and a man may not be a fool only because he makes this mistake, but for other reasons too.

(ii) This argument is unsound, because a wise man may be able to argue on a subject which he does not understand without giving himself away, while a fool could not.

(iii) It is not logical to conclude that a man is a fool because he acts like one in this one particular instance.

(iv) Although fools always make this mistake, it is not stated that all who make this mistake are fools, so that others who are not fools may do so too.

Yes. No (i)..... (ii)..... (iii)..... (iv).....

5. "This pamphlet contains seditious doctrines. The spread of seditious doctrines is dangerous to the State. Therefore this pamphlet must be suppressed".

(i) It is not stated that everything dangerous to the State must be suppressed, so in the premisses given there is no reason for the suppression; and in any case it is not stated that the pamphlet would spread seditious doctrine.

(ii) The spread of seditious doctrines is not always dangerous to the State, for if the State is stable seditious doctrines will not affect it.

(iii) The conclusion is incorrect, for the doctrines in the pamphlet may only appear to be seditious in the opinion of some people. Others may not consider them so.

(iv) To suppress the pamphlet may not of itself avert the danger. The doctrines expressed in it can still be spread verbally by their originators, so other measures may also be necessary.

Yes. No. (i)..... (ii)..... (iii)..... (iv).....

6. "No schoolboy can be expected to understand Constitutional History, and none but schoolboys can be expected to remember dates; so that no one can be expected both to remember dates and to understand Constitutional History".

(i) We cannot assume that this conclusion is true. College students may easily do both, for they are sufficiently developed intellectually to understand Constitutional History, while they are not old enough to have forgotten the dates they learned at school.

(ii) One cannot say that no schoolboy can be expected to understand Constitutional History. A boy who is intelligent and well taught may easily do so.

(iii) Schoolboys should not be expected to remember dates exactly, for this an unsound method of teaching History, so that the whole argument is invalidated.

(iv) The premisses are incomplete. No mention is made of schoolgirls, who are able to remember dates as well as boys.

Yes. No. (i)..... (ii)..... (iii)..... (iv).....

7. "None but those who are contented with their lot in life can justly be considered happy. But the truly wise man will always make himself contented with his lot in life, and, therefore, it follows that he may justly be considered happy".

(i) A wise man can force himself to be contented with his lot in life, but the very fact of this compulsion will prevent his being truly happy.

(ii) Those who are not content with their lot in life are often happy, for there is often more happiness in striving to attain one's desire than in the actual attainment.

(iii) The fact that only those who are contented with their lot in life can justly be considered happy, does not imply that all those who are contented with their lot are of necessity happy.

(iv) The conclusion may be true or not; it will depend on the standard of happiness. Contentment and happiness are not the same thing.

Yes. No. (i).....(ii).....(iii).....(iv).....

8. "In Tutland only Conservatives - and not all of them - are Protectionists (i.e. against free trade): only Liberals - and not all of them - are Home Rulers: but both parties (Conservatives and Liberals) contain supporters of Women's Franchise".

It may be assumed that:
1. No Liberal is a Conservative;
2. all who are not Protectionists are Free Traders;
3. all who do not support Home Rule are Unionists.

This is all that is known about the views of the Tutlanders.

Hence, with this information, it would be wrong to conclude that both Unionists and Free Traders are to be found among the supporters of Women's Franchise.

(i) The Liberals who support Women's Franchise may be those who are Home Rulers, and the Conservatives just those who are Protectionists. Thus there may be neither Free Traders

nor Unionists among the supporters of Women's Franchise.

(ii) Since all Liberals are Free Traders and all Conservatives are Unionists, then there must be both Unionists and Free Traders among those who support Women's Franchise.

(iii) The Liberals who support Women's Franchise may all be Free Traders, and the Conservatives who support Women's Franchise may be those who are not Protectionists, so that these supporters may be all Free Traders.

(iv) The question of Women's Franchise is one which is not affected by considerations of Unionism or Free Trade, so these considerations are irrelevant.

Yes. It <u>would</u> be wrong.
No. It would <u>not</u> be wrong:
(i).....(ii).....(iii).....(iv).....